Evaluation and Combination of Conditional Quantile Forecasts

Raffaella GIACOMINI
Department of Economics, University of California, Los Angeles, P.O. Box 951477, Los Angeles, CA 90095-1477 (*giacomin@econ.ucla.edu*)

Ivana KOMUNJER
Division of the Humanities and Social Sciences, California Institute of Technology, MC 228-77, Pasadena, CA 91125 (*komunjer@hss.caltech.edu*)

We propose an encompassing test for comparing conditional quantile forecasts in an out-of-sample framework. Our test provides a basis for forecast combination when encompassing is rejected. Its central features are (1) use of the "tick" loss function, (2) a conditional approach to out-of-sample evaluation, and (3) derivation in an environment with asymptotically nonvanishing estimation uncertainty. Our approach is valid under general conditions; the forecasts can be based on nested or nonnested models and can be obtained by general estimation procedures. We illustrate the test properties in a Monte Carlo experiment and apply it to evaluate and compare four popular value-at-risk models.

KEY WORDS: Encompassing; Generalized method of moments; Tick loss function; Value-at-risk.

1. INTRODUCTION

The vast majority of the economic forecasting literature has traditionally focused on producing and evaluating point forecasts for the conditional mean of some variable of interest. More recently, increasing attention has been devoted to other characteristics of the unknown forecast distribution, besides its conditional mean, such as a particular conditional quantile.

A primary example of the growing interest in conditional quantile forecasts is in the context of risk management, as witnessed by the literature on value-at-risk (VaR) (e.g., Duffie and Pan 1997). Ever since August 1996, when U.S. bank regulators adopted a "market risk" supplement to the Basle Accord (1988), the regulatory capital requirements of commercial banks with trading activities are based on VaR estimates. This important measure of market risk is defined as the opposite of a prespecified quantile of the conditional distribution of portfolio returns, and its estimates are routinely generated by the banks' internal models. There are a variety of approaches to estimating conditional quantiles in general and VaR in particular, ranging from parametric (e.g., Danielsson and de Vries 1997; Barone-Adesi, Bourgoin, and Giannopoulos 1998; Diebold, Schuermann, and Stroughair 1998; Embrechts, Resnick, and Samorodnitsky 1999; McNeil and Frey 2000), to semiparametric (e.g., Koenker and Zhao 1996; Taylor 1999; Chernozhukov and Umanstev 2001; Christoffersen, Hahn, and Inoue 2001; Engle and Manganelli 2004; Komunjer 2005), to nonparametric (e.g., Battacharya and Gangopadhyay 1990; White 1992).

Given the range of techniques available for producing conditional quantile forecasts, it is necessary to develop adequate tools for their evaluation. A number of authors have focused on *absolute* evaluation, that is, on testing whether a forecasting model is correctly specified or whether a sequence of forecasts satisfies certain optimality properties. For example, Zheng (1998) and Bierens and Ginther (2001) proposed specification tests for evaluating a model against a generic alternative. Christoffersen (1998) proposed a "correct conditional coverage" criterion for evaluating a sequence of interval forecasts that does not require knowledge of the underlying model. Corradi and Swanson (2002) allowed for misspecification and proposed a test that compares a reference model against generic nonlinear alternatives. A potential problem with absolute evaluation is that if different models are rejected as being misspecified or if they are all accepted, then we are left without any guidance as to which one to choose. The approach of Corradi and Swanson (2002) is similarly inconclusive if the reference model is rejected. In this article, we thus focus on *relative* evaluation, which involves comparing the performance of competing, possibly misspecified models or sequences of forecasts for a variable and choosing the one that performs the best. This approach was taken by Christoffersen et al. (2001), who proposed a method for comparing nonnested VaR estimates. These authors assumed that the VaR is a linear function of the volatility and proposed estimating the parameters by the information theoretic alternative to generalized method of moments (GMM) due to Kitamura and Stutzer (1997). The evaluation of Christoffersen et al. (2001) was conducted in-sample and is valid only if the returns belong to a location-scale family (which implies that the VaR is a linear function of the volatility). Further, to apply their test, all VaR forecasts must be obtained by the estimation method of Kitamura and Stutzer (1997).

It is frequently the case, however, that good in-sample performance does not imply good out-of-sample performance and that the models underlying the forecasts remain partially or completely unknown to the forecast user. Moreover, given the variety of approaches to estimating conditional quantile models outlined earlier, it may be of interest to investigate whether different estimation techniques have an effect on forecast performance. In general, when several forecasts of the same variable are available, it is desirable to have formal testing procedures for out-of-sample comparison that do not necessarily require knowledge of the underlying model or, if the model is known,

Journal of Business & Economic Statistics
October 2005, Vol. 23, No. 4
DOI 10.1198/073500105000000018

do not restrict attention to a specific estimation procedure. The goal of this article is to provide such a test.

Given an appropriate choice of loss function, one could in principle compare the out-of-sample average loss implied by alternative quantile forecasts using the tests of equal predictive ability proposed by Diebold and Mariano (1995), West (1996), White (2000), and Corradi and Swanson (2004). But these approaches may not be applicable in several important cases, such as when the forecasts are from nested models or when they depend on semiparametric or nonparametric estimators.

In this article we choose a different approach and construct a test for out-of-sample conditional quantile forecast comparison based on the principle of encompassing (e.g., Hendry and Richard 1982; Mizon and Richard 1986).

The novel features of our implementation of the principle of encompassing are first, the choice of the relevant loss function, which we argue to be the "tick" loss function; second, the focus on *conditional* expected loss rather than unconditional expected loss in the formulation of the encompassing test; and third, the derivation of our test in an environment with asymptotically nonvanishing estimation uncertainty. These last two features link the approach in this article to that of Giacomini and White (2003), who proposed a general framework for out-of-sample predictive ability testing. Some of the advantages of this framework over that of, for instance, West (1996), are that it allows the forecasts to be generated by parametric models as well as by semiparametric or nonparametric techniques, and that it is applicable to both nested and nonnested forecast comparisons. The implementation of our test makes use of standard GMM techniques. As a byproduct, our framework also provides a link to Christoffersen's (1998) "correct conditional coverage" criterion for the absolute evaluation of interval forecasts.

A final feature of our encompassing approach is that it gives a theoretical basis for quantile forecast combination in cases when neither forecast encompasses its competitor. From a theoretical viewpoint, forecast combination can be seen as a way to pool the information contained in the individual forecasts, and its benefits have been widely advocated since the early work of Bates and Granger (1969). According to Granger (1989), there are two situations when it is useful to combine forecasts. If the forecasts are based on the same information set, then a forecast combination can be useful only if the original forecasts are suboptimal according to the relevant loss function. But if the forecasts are instead based on different information sets, then combining them can potentially improve the forecasting performance by pooling the information contained in the two sets. Recent empirical work by Stock and Watson (1999, 2003) has further confirmed the accuracy gains induced by forecast combination for a large number of macroeconomic and financial time series. Surprisingly little empirical work has been done in the context of conditional quantile forecasting. Yet the benefits of expanding the information set through combination might be particularly evident for quantiles with small nominal coverage, as is usually the case for VaR. Extreme quantiles are very sensitive to the few observations in the tails of the empirical distribution of the sample, and combining forecasts based on different information sets can thus be seen as a way to make the forecast performance more robust to the effects of sample-specific factors.

We illustrate the usefulness of the conditional quantile forecast encompassing (CQFE) test by applying it to the problem of VaR evaluation using S&P500 daily return data. We consider popular models for producing 1% and 5% VaR forecasts and generally conclude that the forecast combination outperforms the individual forecasts.

The remainder of the article is organized as follows. Section 2 describes the environment and gives an overview of our encompassing approach to comparing and combining competing conditional quantile forecasts. Section 3 introduces the test for conditional quantile forecast encompassing and discusses the estimation problem underlying implementation of the test. A formal definition of the CQFE test statistic is provided in Theorem 1 in Section 3.2, which is the main result of this article. Section 4 analyzes the small-sample size and power properties of the proposed test, and Section 5 applies the test to the problem of VaR forecast evaluation and combination. Section 6 concludes. The Appendix presents proofs.

2. OVERVIEW

2.1 Description of the Environment

Consider a stochastic process $\mathbf{X} \equiv \{\mathbf{X}_t : \Omega \longrightarrow \mathbb{R}^{k+1},\ k \in \mathbb{N},\ t = 1, \ldots, T\}$ defined on a complete probability space $(\Omega, \mathcal{F}, P)$, where $\mathcal{F} \equiv \{\mathcal{F}_t,\ t = 1, \ldots, T\}$ and $\mathcal{F}_t \equiv \sigma\{\mathbf{X}_s,\ s \leq t\}$. We partition the observed vector $\mathbf{X}_t$ as $\mathbf{X}_t \equiv (Y_t, \mathbf{Z}_t')'$, where $Y_t : \Omega \to \mathbb{R}$ is a continuous random variable of interest and $\mathbf{Z}_t : \Omega \to \mathbb{R}^k$ is a vector of explanatory variables. We are interested in the α-quantile of the distribution of Y_{t+1} conditional on the information set $\mathcal{F}_t$, $Q_{t,\alpha}$, defined as

$$P_t(Y_{t+1} \leq Q_{t,\alpha}) = \alpha \tag{1}$$

or

$$Q_{t,\alpha} \equiv F_t^{-1}(\alpha), \tag{2}$$

where $\alpha \in (0, 1)$, F_t is the conditional distribution function of Y_{t+1} and F_t^{-1} its inverse. Using the standard convention, the subscript t under the probability $P(\cdot)$, distribution function $F(\cdot)$, density $f(\cdot)$, expectation $E[\cdot]$ or α-quantile Q denotes conditioning on the information set $\mathcal{F}_t$. To further simplify the notation, we hereafter drop the reference to the index α and simply denote the time t conditional α-quantile by Q_t. As a general rule, a lower-case letter is used to denote observations of the corresponding random variable (e.g., $\mathbf{x}_t$ and $\mathbf{X}_t$).

Our goal is to propose a test for comparing alternative sequences of one-step-ahead forecasts of Q_t. We perform the evaluation in an out-of-sample fashion. This involves dividing the sample of size T into an in-sample part of size m and an out-of-sample part of size n, so that $T = m + n$. The in-sample portion is used to produce the first set of forecasts, and the evaluation is performed over the remaining out-of-sample portion. We impose few restrictions on how the forecasts are produced. In particular, they may be based on parametric models or be generated by semiparametric or nonparametric techniques. The forecasts can be produced using either a fixed forecasting scheme or a rolling window forecasting scheme. For example, for a parametric model, a fixed forecasting scheme involves estimating the parameters only once on the first m observations and using

these estimates to produce all of the forecasts for the out-of-sample period $t = m + 1, \ldots, T$. A rolling window forecasting scheme, in contrast, implies reestimating the parameters at each out-of-sample point t using an estimation sample containing the m most recent observations, that is, the observations from $t - m + 1$ to t.

Let $\hat{\boldsymbol{\beta}}_{t,m}$ denote the $k \times 1$ vector collecting the time-t estimated parameters from the two models (for parametric forecasting) or whatever semiparametric or nonparametric estimator used in the construction of the forecasts. In what follows, we use the common notation $\hat{\boldsymbol{\beta}}_{t,m}$ for either forecasting scheme, with the understanding that a fixed forecasting scheme corresponds to the case where $\hat{\boldsymbol{\beta}}_{t,m} = \hat{\boldsymbol{\beta}}_{m,m}$ for all t, $m \le t \le T - 1$, whereas for the rolling window forecasting scheme, $\hat{\boldsymbol{\beta}}_{t,m}$ changes with t but depends only on the previous m observations.

For simplicity, we restrict attention to pairwise comparisons, but all of the techniques can be readily applied to the case of multiple forecasts. For each time t, $m \le t \le T - 1$, the one-step-ahead forecasts of Q_t formulated at time t are denoted by $\hat{q}_{1m,t} \equiv q_1(x_t, x_{t-1}, \ldots; \hat{\boldsymbol{\beta}}_{t,m})$ and $\hat{q}_{2m,t} \equiv q_2(x_t, x_{t-1}, \ldots; \hat{\boldsymbol{\beta}}_{t,m})$, where q_1 and q_2 are $\mathcal{F}_t$-measurable functions.

The crucial requirement that we impose on the functions q_1 and q_2 is that they are constant over time. This implies, in particular, that use of an expanding estimation window (recursive forecasting scheme) is not allowed, whereas forecasting schemes using either a fixed or a rolling window of constant length satisfy the requirement. In the remainder of the article, we assume that the in-sample size m is a finite constant, chosen by the user a priori. As a consequence, all of our results should be interpreted as being conditional on the given choice of m, but for ease of notation we choose not to make this dependence explicit (except for $\hat{q}_{1m,t}$ and $\hat{q}_{2m,t}$).

2.2 Principles of Forecast Encompassing

Our approach to comparing conditional quantile forecasts is based on the principle of encompassing. Following, for example, Hendry and Richard (1982), Mizon and Richard (1986), and Diebold (1989), encompassing arises when one of two competing forecasts is able to explain the predictive ability of its rival. According to Clements and Hendry (1998, p. 228), a test for forecast encompassing can be generally defined as follows:

> A test for forecast encompassing is a test of the conditional efficiency of a forecast, where a forecast is said to be conditionally efficient if the expected loss of a combination of that forecast and a rival forecast is not significantly less than the expected loss of the original forecast alone.

The two key ingredients of any forecast encompassing test are, therefore (1) the loss function involved in the computation of the expected loss and (2) the weights of the forecast combination. The choice of the loss function is closely related to which characteristic of the unknown future distribution of the variable one wants to forecast. Let $\hat{f}_t$ be a forecast of some characteristic of interest of the random variable Y_{t+1}, conditional on the information set at time t. The forecast $\hat{f}_t$ is said to be optimal at time t if it minimizes $E_t[L(Y_{t+1} - \hat{f}_t)]$, where L is some loss function, $L : \mathbb{R} \to \mathbb{R}^+$. Note that the optimal forecast minimizes the expected loss *conditional* on $\mathcal{F}_t$. As discussed in detail later, the focus on conditional (rather than unconditional) expected loss is a central feature of our treatment of both evaluation and combination of forecasts and distinguishes our approach from related literature (e.g., Granger 1989; Taylor and Bunn 1998; Elliott and Timmermann 2004).

Different loss functions L correspond to different optimal forecasts. For example, letting $e_{t+1} \equiv y_{t+1} - \hat{f}_t$, if a quadratic loss function $L(e_{t+1}) = e_{t+1}^2$ is used, then the optimal forecast is the conditional mean of the distribution of Y_{t+1}. If, on the other hand, an absolute value loss function $L(e_{t+1}) = |e_{t+1}|$ is used, then the optimal forecast corresponds to the conditional median of the distribution of Y_{t+1}. In this article the object of interest is Q_t, the conditional α-quantile of the distribution of Y_{t+1}. The corresponding loss function, L, is the asymmetric linear loss function of order α, $\mathcal{T}_\alpha$, defined as

$$\mathcal{T}_\alpha(e_{t+1}) \equiv \big(\alpha - \mathbb{1}(e_{t+1} < 0)\big)e_{t+1}, \tag{3}$$

which is also known as the "tick" or "check" loss function in the literature. We can thus argue that the tick function $\mathcal{T}$ is the implicit loss function whenever the object of interest is a forecast of a particular quantile of the conditional distribution of Y_{t+1}.

Regarding the choice of weights in the combination, in this article we restrict attention to linear combinations, $(\theta_{1t} \times \hat{q}_{1m,t} + \theta_{2t}\hat{q}_{2m,t})$, where $(\theta_{1t}, \theta_{2t})$ lies in some compact subset of $\mathbb{R}^2$. The values of θ_{1t} and θ_{2t} can be further constrained to lie in $(0, 1)$, with $\theta_{1t} + \theta_{2t} = 1$, but we choose not to impose this restriction herein. (For a discussion of restrictions on the combination weights, see, e.g., Granger and Ramanathan 1984.) In the next section we formalize the concept of encompassing for conditional quantile forecasts.

2.3 Encompassing for Conditional Quantiles

Based on the general idea of Clements and Hendry (1998), we say that forecast $\hat{q}_{1m,t}$ encompasses forecast $\hat{q}_{2m,t}$ at time t if and only if

$$E_t[\mathcal{T}_\alpha(Y_{t+1} - \hat{q}_{1m,t})] \le E_t[\mathcal{T}_\alpha(Y_{t+1} - (\theta_{1t}\hat{q}_{1m,t} + \theta_{2t}\hat{q}_{2m,t})] \quad \text{a.s.-}P, \text{ for all } (\theta_{1t}, \theta_{2t}) \in \Theta, \tag{4}$$

where Θ is a compact subset of $\mathbb{R}^2$. In practice, testing the inequality (4) is not feasible, because it involves computing the expected loss for all $(\theta_{1t}, \theta_{2t}) \in \Theta$. Instead, let $(\theta_{1t}^*, \theta_{2t}^*)$ denote the optimal set of weights, defined as a solution to the minimization problem $\min_{(\theta_1,\theta_2)\in\Theta} E_t[\mathcal{T}_\alpha(Y_{t+1} - (\theta_1 \times \hat{q}_{1m,t} + \theta_2\hat{q}_{2m,t}))]$. We then have that $E_t[\mathcal{T}_\alpha(Y_{t+1} - (\theta_{1t}^*\hat{q}_{1m,t} + \theta_{2t}^*\hat{q}_{2m,t}))] \le E_t[\mathcal{T}_\alpha(Y_{t+1} - (\theta_{1t}\hat{q}_{1m,t} + \theta_{2t}\hat{q}_{2m,t})]$, for every $(\theta_{1t}, \theta_{2t}) \in \Theta$, which implies that $E_t[\mathcal{T}_\alpha(Y_{t+1} - (\theta_{1t}^*\hat{q}_{1m,t} + \theta_{2t}^*\hat{q}_{2m,t}))] \le E_t[\mathcal{T}_\alpha(Y_{t+1} - \hat{q}_{1m,t})]$. Hence we obtain the following definition of encompassing.

Definition 1 (Conditional quantile forecast encompassing). Let $\hat{q}_{1m,t}$ and $\hat{q}_{2m,t}$ be alternative forecasts for Q_t. $\hat{q}_{1m,t}$ is said to encompass $\hat{q}_{2m,t}$ at time t if and only if

$$E_t[\mathcal{T}_\alpha(Y_{t+1} - \hat{q}_{1m,t})] = E_t[\mathcal{T}_\alpha(Y_{t+1} - (\theta_{1t}^*\hat{q}_{1m,t} + \theta_{2t}^*\hat{q}_{2m,t})], \quad \text{a.s.-}P, \tag{5}$$

that is, if and only if

$$(\theta_{1t}^*, \theta_{2t}^*) = (1, 0), \tag{6}$$

where $\mathcal{T}_\alpha$ is the tick loss function defined in (3) and $(\theta_{1t}^*, \theta_{2t}^*)$ is such that

$$(\theta_{1t}^*, \theta_{2t}^*) \equiv \arg \min_{(\theta_1,\theta_2)\in\Theta} E_t\big[\mathcal{T}_\alpha\big(Y_{t+1} - (\theta_1 \hat{q}_{1m,t} + \theta_2 \hat{q}_{2m,t})\big)\big]. \tag{7}$$

The equivalence between (5) and (6) follows from the fact that the right side of (5) is the minimum of the conditional expected loss over Θ.

Comments.

1. If we interpret a conditional expectation as a prediction, then equality (5) can be viewed as saying that $\hat{q}_{1m,t}$ encompasses $\hat{q}_{2m,t}$ if the forecaster cannot predict whether the optimal combination of the two forecasts will outperform the original forecast at time $t+1$, given what is known at time t. This focus on prediction of future performance (conditional expectation) rather than on assessment of average performance (unconditional expectation) in the definition of encompassing distinguishes our approach from the classic encompassing literature (e.g., Hendry and Richard 1982; Mizon and Richard 1986) and establishes a link with the general framework for predictive ability testing proposed by Giacomini and White (2003).

2. Similar to Giacomini and White (2003), the forecasts $\hat{q}_{1m,t}$ and $\hat{q}_{2m,t}$ in our definition of encompassing depend on the parameter estimates at time t, rather than on population values as in, for example, the approach of West (2001). This corresponds to a shift from evaluating a forecast model to evaluating the "forecast method," which includes the model as well as the estimation procedure and the choice of estimation window.

3. Focusing on the actual forecasts rather than the underlying models in the definition of encompassing means that we do not assume that the forecasts are estimated using the same tick loss function used for the evaluation. As a result, we provide a unified framework for comparing forecasts obtained by possibly different estimation techniques.

In the next section we discuss implementation of the CQFE test.

3. CONDITIONAL QUANTILE FORECAST ENCOMPASSING TEST

In what follows, we are interested in testing whether $\hat{q}_{1m,t}$ conditionally encompasses $\hat{q}_{2m,t}$ over the entire out-of-sample period $t = m, \ldots, T-1$. Hereafter, we let $\boldsymbol{\theta} \equiv (\theta_1, \theta_2)'$ and $\hat{\mathbf{q}}_{m,t} \equiv (\hat{q}_{1m,t}, \hat{q}_{2m,t})'$. The following lemma expresses the optimal weights in terms of the optimization problem's first-order condition.

Lemma 1 (Correct conditional coverage criterion). The vector of optimal weights $\boldsymbol{\theta}_t^*$ defined in (7) satisfies the following first-order condition:

$$E_t[\alpha - \mathbb{1}(Y_{t+1} - \boldsymbol{\theta}_t^{*\prime}\hat{\mathbf{q}}_{m,t} < 0)] = 0, \quad \text{a.s.-}P. \tag{8}$$

It is interesting to note that (8) corresponds exactly to Christoffersen's (1998) "correct conditional coverage criterion" for evaluating interval forecasts, here applied to the combination forecast $\boldsymbol{\theta}_t^{*\prime}\hat{\mathbf{q}}_{m,t}$. What Lemma 1 shows is that the correct conditional coverage condition is equivalent to requiring optimality of an interval forecast with respect to the tick loss function.

We now discuss estimation of the optimal combination weights.

3.1 Generalized Method-of-Moments Estimation of Optimal Combination Weights

According to Definition 1, $\hat{q}_{1m,t}$ conditionally encompasses $\hat{q}_{2m,t}$ for all t, $m \le t \le T-1$ if and only if $\boldsymbol{\theta}_m^* = \cdots = \boldsymbol{\theta}_{T-1}^* = (1,0)'$. In other words, the optimal combination weights are constant in time and equal to $(1,0)'$. By Lemma 1, it should therefore be the case that for $\mathbf{e}_1 = (1,0)'$, $E[(\alpha - \mathbb{1}(Y_{t+1} - \mathbf{e}_1'\hat{\mathbf{q}}_{m,t} < 0))\mathbf{W}_t] = 0$, for all $\mathcal{F}_t$-measurable functions $\mathbf{W}_t$ and for all t, $m \le t \le T-1$. Let $\mathbf{W}_t^*$ be an $h \times 1$ vector of variables that are observed at time t and that contain all of the relevant information from $\mathcal{F}_t$. We refer to $\mathbf{W}_t^*$ as the "information vector." As stated in Proposition 1, the general requirement on $\{\mathbf{W}_t^*\}$ is that it is a strictly stationary and mixing series. As such, we allow $\mathbf{W}_t^*$ to include previous forecasts (or measures of past forecast performance), provided that they are produced by either a fixed or a rolling window forecasting scheme. The reason for this is that in these two cases the forecasts are constant measurable functions of a finite window of data and thus inherit the properties of stationarity and mixing from the underlying series. In practice, the choice of $\mathbf{W}_t^*$ depends on the nature of the application considered, as we discuss in more detail in Section 5. Further, denote by $\mathbf{g}$ an h-vector-valued function $\mathbf{g}: \Theta \times \mathbb{R} \times \mathbb{R}^h \to \mathbb{R}^h$ such that

$$\mathbf{g}(\boldsymbol{\theta}; y_{t+1}, \mathbf{w}_t^*) \equiv [\alpha - \mathbb{1}(y_{t+1} - \boldsymbol{\theta}'\hat{\mathbf{q}}_{m,t} < 0)]\mathbf{w}_t^*. \tag{9}$$

The key element in our implementation of the encompassing test is that under the null of encompassing, we have $E[\mathbf{g}(\mathbf{e}_1; Y_{t+1}, \mathbf{W}_t^*)] = \mathbf{0}$, and hence we can use Hansen's (1982) GMM approach to estimate the solution $\boldsymbol{\theta}^*$ to the moment condition

$$\mathbf{g}_0(\boldsymbol{\theta}^*) \equiv E[\mathbf{g}(\boldsymbol{\theta}^*; Y_{t+1}, \mathbf{W}_t^*)] = \mathbf{0}, \tag{10}$$

and then test whether $\boldsymbol{\theta}^* = \mathbf{e}_1$. Given the out-of-sample portion of size $n = T - m$, consisting of the sequence of observations $(\mathbf{w}_m^{*\prime}, y_{m+1}, \ldots, \mathbf{w}_{T-1}^{*\prime}, y_T)'$, the GMM estimator of $\boldsymbol{\theta}^*$, denoted by $\hat{\boldsymbol{\theta}}_n$, is defined as a solution to the minimization problem

$$\min_{\boldsymbol{\theta}\in\Theta} [\mathbf{g}_n(\boldsymbol{\theta})]'\hat{\mathbf{S}}_n^{-1}[\mathbf{g}_n(\boldsymbol{\theta})], \tag{11}$$

where $\mathbf{g}_n(\cdot)$ is the sample moment function, $\mathbf{g}_n(\boldsymbol{\theta}) \equiv n^{-1} \times \sum_{t=m}^{T-1} \mathbf{g}(\boldsymbol{\theta}; y_{t+1}, \mathbf{w}_t^*)$, and $\hat{\mathbf{S}}_n$ is a consistent estimator of the asymptotic variance matrix $\mathbf{S}$,

$$\mathbf{S} \equiv E[\mathbf{g}(\boldsymbol{\theta}^*; Y_{t+1}, \mathbf{W}_t^*)\mathbf{g}(\boldsymbol{\theta}^*; Y_{t+1}, \mathbf{W}_t^*)']. \tag{12}$$

Using the fact that the first-order condition (8) implies that $\{\mathbf{g}(\boldsymbol{\theta}^*; Y_{t+1}, \mathbf{W}_t^*), \mathcal{F}_t\}$ is a martingale difference sequence, a consistent estimator of $\mathbf{S}$ is given by

$$\begin{aligned}\hat{\mathbf{S}}_n(\tilde{\boldsymbol{\theta}}_n) &\equiv n^{-1}\sum_{t=m}^{T-1} \mathbf{g}(\tilde{\boldsymbol{\theta}}_n; y_{t+1}, \mathbf{w}_t^*)\mathbf{g}(\tilde{\boldsymbol{\theta}}_n; y_{t+1}, \mathbf{w}_t^*)' \\ &= n^{-1}\sum_{t=m}^{T-1} [\alpha - \mathbb{1}(y_{t+1} - \tilde{\boldsymbol{\theta}}_n'\hat{\mathbf{q}}_{m,t} < 0)]^2 \mathbf{w}_t^*\mathbf{w}_t^{*\prime}, \end{aligned} \tag{13}$$

where $\tilde{\boldsymbol{\theta}}_n$ is some initial consistent estimate of $\boldsymbol{\theta}^*$. In cases when the information vector fails to incorporate all of the relevant information, condition $\mathbf{g}_0(\boldsymbol{\theta}^*) = \mathbf{0}$ is no longer equivalent to the first-order condition (8) and $\{\mathbf{g}(\boldsymbol{\theta}^*; Y_{t+1}, \mathbf{W}_t^*)\}$ is no longer a martingale difference sequence. However, $\mathbf{S}$ can still be consistently estimated using some heteroscedasticity- and autocorrelation-robust estimator, like Newey and West's (1987) estimator. We now focus on the asymptotic properties of the GMM estimator $\hat{\boldsymbol{\theta}}_n$.

Proposition 1 (Consistency). Assume that for every t, $m \leq t \leq T-1$, (a) the conditional density of Y_{t+1}, $f_t(\cdot)$, is continuous and strictly positive; (b) for $i = 1, 2$, $\hat{q}_{im,t} \neq 0$, a.s.-P, and $\operatorname{corr}(\hat{q}_{1m,t}, \hat{q}_{2m,t}) \neq \pm 1$; (c) $\{(\mathbf{W}_t^{*\prime}, \mathbf{X}_t')'\}$ is strictly stationary and α-mixing with α of size $-r/(r-2)$, with $r > 2$; (d) $E[\mathbf{W}_t^* \mathbf{W}_t^{*\prime}]$is nonsingular; and (e) there exist some $\delta > 0$ such that $E\|\mathbf{W}_t^*\|^{2r+\delta} < \infty$. Then $\hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}^*$, as $n \to \infty$.

Assumption (b) is a mild condition ruling out the possibility that the two sequences of forecasts are perfectly correlated, which would happen if, for example, the two models were proportional or differed only by a constant. One could in principle relax the assumption of strict stationarity in (c) and rely on existing results on the consistency and asymptotic normality of GMM estimators for mixing sequences. However, we decided not to pursue this option, because it would cause the optimal weights to depend on the sample size, and thus result in a less intuitive formulation of the null hypothesis of encompassing. Conditions (d) and (e) are fairly standard and imply in particular that all of the components of the information vector are not linearly dependent.

We now turn to the asymptotic distribution of $\hat{\boldsymbol{\theta}}_n$. The standard asymptotic normality results for GMM require that $\mathbf{g}_n(\boldsymbol{\theta})$ be once differentiable, which is not the case here. There are, however, asymptotic normality results for nonsmooth moment functions, hereinafter we use the one proposed by Newey and McFadden (1994). The basic insight of their approach is that a smoothness condition on $\mathbf{g}_n(\boldsymbol{\theta})$ can be replaced by the smoothness of its limit $\mathbf{g}_0(\boldsymbol{\theta})$, with the requirement that certain remainder terms are small. The asymptotic distribution of $\hat{\boldsymbol{\theta}}_n$ is derived in the next proposition.

Proposition 2 (Asymptotic normality). Let the assumptions of Proposition 1 hold and further assume that (f) $E\|\hat{\mathbf{q}}_{m,t}\|^4 < \infty$; (g) the conditional density of Y_{t+1}, $f_t(\cdot)$, is bounded; and (h) $\boldsymbol{\theta}^*$ is an interior point of Θ. Then $\hat{\boldsymbol{\theta}}_n$ is asymptotically normal, $(\boldsymbol{\gamma}'\mathbf{S}^{-1}\boldsymbol{\gamma})^{-1/2}\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) \xrightarrow{d} \mathcal{N}(0, 1)$, with

$$\boldsymbol{\gamma} \equiv -E\big[f_t(\boldsymbol{\theta}^{*\prime}\hat{\mathbf{q}}_{m,t})\mathbf{W}_t^*\hat{\mathbf{q}}_{m,t}'\big], \tag{14}$$

and $\mathbf{S}$ as defined in (12).

Note that the expression for $\boldsymbol{\gamma}_n$, which depends on the value of the conditional density f_t evaluated at the optimal combination of quantiles, is similar to the one commonly found in the quantile regression literature (e.g., Koenker and Bassett 1978; Komunjer 2005). Further, note that assumption (f) implicitly places conditions on the existence of the finite-sample moments of the estimator on which $\hat{\mathbf{q}}_{m,t}$ is based.

3.2 CQFE Test Statistic

We consider conducting two separate tests: $H_{10}: (\theta_1^*, \theta_2^*) = (1, 0)$ against $H_{1a}: (\theta_1^*, \theta_2^*) \neq (1, 0)$, and $H_{20}: (\theta_1^*, \theta_2^*) = (0, 1)$ against $H_{2a}: (\theta_1^*, \theta_2^*) \neq (0, 1)$, which correspond to testing whether forecast $\hat{q}_{1m,t}$ encompasses $\hat{q}_{2m,t}$ or whether $\hat{q}_{2m,t}$ encompasses $\hat{q}_{1m,t}$. We propose a Wald test of hypotheses H_{10} and H_{20} in the following theorem.

Theorem 1 (CQFE test). Let the assumptions of Proposition 2 hold. Consider the test statistics

$$ENC1_n = n\big((\hat{\theta}_{1n}, \hat{\theta}_{2n}) - (1, 0)\big)\hat{\boldsymbol{\Omega}}_n^{-1}\big((\hat{\theta}_{1n}, \hat{\theta}_{2n}) - (1, 0)\big)' \tag{15}$$

and

$$ENC2_n = n\big((\hat{\theta}_{1n}, \hat{\theta}_{2n}) - (0, 1)\big) \times \hat{\boldsymbol{\Omega}}_n^{-1}\big((\hat{\theta}_{1n}, \hat{\theta}_{2n}) - (0, 1)\big)', \tag{16}$$

where $(\hat{\theta}_{1n}, \hat{\theta}_{2n})$ are defined in (11) and $\hat{\boldsymbol{\Omega}}_n$ is some consistent estimate of $\boldsymbol{\Omega} \equiv (\boldsymbol{\gamma}'\mathbf{S}^{-1}\boldsymbol{\gamma})^{-1}$. Then (a) under H_{i0}: $ENCk_n \xrightarrow{d} \chi_2^2$, as $n \to \infty$, $k = 1, 2$, and (b) under H_{ia}: $ENCk_n \to +\infty$, as $n \to \infty$, $k = 1, 2$.

Similar to the approach of Giacomini and White (2003), and in contrast to the existing predictive ability testing literature (e.g., West 1996; McCracken 2000), our asymptotic framework lets the number of out-of-sample observations go to infinity, while the in-sample size m remains finite. We adopt this assumption as a convenient way to obtain an environment with nonvanishing estimation uncertainty, which results in our test having several advantages. It can directly capture the effect of estimation uncertainty on forecast performance, it allows for general estimation procedure, and it can avoid the problems associated with comparison of predictive ability involving nested models. To see why this is the case, suppose that we were comparing nested models and that the smaller model were correctly specified. Letting the size of the estimation window go to infinity would cause the parameter estimates to converge to their probability limits, which would render the forecasts from the two models asymptotically perfectly correlated, thereby invalidating assumption (b) of Proposition 1.

Note that the assumption that m is finite rules out using an expanding estimation window forecasting scheme. As noted by a referee, a drawback of requiring that observations from the distant past be dropped from the estimation sample is that this may result in suboptimal parameter estimates in a stationary environment. In principle, one could create an environment with nonvanishing estimation uncertainty in the context of an expanding estimation window forecasting scheme by assuming that the in-sample size grows more slowly than the out-of-sample size, but we decided against imposing this artificial condition here.

3.3 Test Implementation and Forecast Selection Implications

In the computation of the test statistics $ENC1_n$ and $ENC2_n$, defined in Theorem 1, we need a consistent estimator of the asymptotic covariance matrix $\boldsymbol{\Omega} = (\boldsymbol{\gamma}'\mathbf{S}^{-1}\boldsymbol{\gamma})^{-1}$ derived in Proposition 2. We estimate $\mathbf{S}$ using the sample variance of our moment vector $\mathbf{g}$, $\hat{\mathbf{S}}_n \equiv \hat{\mathbf{S}}(\hat{\boldsymbol{\theta}}_n)$, which is a fairly commonly used

estimator. The computation of $\hat{\boldsymbol{\theta}}_n$ and $\hat{\mathbf{S}}_n$ is typically done recursively. We first choose an $r \times r$ identity-weighting matrix, $\mathbf{I}_{r\times r}$, in (11) and compute the corresponding $\hat{\theta}_{n,1}$. The resulting new weighting matrix, $\hat{\mathbf{S}}_n^{-1}(\hat{\theta}_{n,1})$, is more efficient than the previous one, and solving (11) leads to a new estimator $\hat{\theta}_{n,2}$. The last two steps can then be repeated until $\hat{\theta}_{n,j}$ equals its previous value, $\hat{\theta}_{n,j-1}$. Unlike $\hat{\mathbf{S}}_n$, our estimator of the matrix $\boldsymbol{\gamma}$ in (14) has a novel form, not yet seen in the literature. We let

$$\hat{\boldsymbol{\gamma}}_{n,\tau} \equiv -n^{-1}\sum_{t=m}^{T-1}\frac{1}{\tau}\exp[(y_{t+1}-\hat{\boldsymbol{\theta}}_n'\hat{\mathbf{q}}_{m,t})/\tau] \times \mathbb{1}(y_{t+1}-\hat{\boldsymbol{\theta}}_n'\hat{\mathbf{q}}_{m,t}<0)\mathbf{w}_t^*\hat{\mathbf{q}}_{m,t}', \quad (17)$$

with $\tau > 0$. The foregoing estimator $\hat{\boldsymbol{\gamma}}_{n,\tau}$ is obtained as a derivative of a smooth approximation $\mathbf{g}_{n,\tau}(\boldsymbol{\theta})$ to the sample moment function $\mathbf{g}_n(\boldsymbol{\theta})$, defined as $\mathbf{g}_{n,\tau}(\boldsymbol{\theta}) \equiv n^{-1}\sum_{t=m}^{T-1}\{\alpha - [1-\exp((y_{t+1}-\boldsymbol{\theta}'\hat{\mathbf{q}}_{m,t})/\tau)]\mathbb{1}(y_{t+1}-\boldsymbol{\theta}'\hat{\mathbf{q}}_{m,t}<0)\}\mathbf{w}_t^*$ (see, e.g., Bracewell 2000, pp. 63–65). As τ goes to 0, the term inside the curly brackets converges to $\alpha - \mathbb{1}(y_{t+1}-\boldsymbol{\theta}'\hat{\mathbf{q}}_{m,t}<0)$ and is hence a smooth approximation to the indicator function. The convergence of $\hat{\boldsymbol{\gamma}}_{n,\tau}$ to its expected value is uniform in τ in a neighborhood of 0, which ensures that $\lim_{\tau\to 0}\hat{\boldsymbol{\gamma}}_{n,\tau} \xrightarrow{p} \boldsymbol{\gamma}$, as we show in the following lemma.

Lemma 2. Under the assumptions of Proposition 2, $\lim_{\tau\to 0}\hat{\boldsymbol{\gamma}}_{n,\tau} \xrightarrow{p} \boldsymbol{\gamma}$ and $\hat{\boldsymbol{\Omega}}_n \equiv \lim_{\tau\to 0}(\hat{\boldsymbol{\gamma}}_{n,\tau}\hat{\mathbf{S}}_n^{-1}\hat{\boldsymbol{\gamma}}_{n,\tau})^{-1} \xrightarrow{p} \boldsymbol{\Omega}$.

The CQFE test can then be implemented as follows. For a desired level of confidence, one first chooses the corresponding critical value c from the χ_2^2 distribution. If $ENC1_n \le c$, then we conclude that $\hat{q}_{1m,t}$ encompasses $\hat{q}_{2m,t}$. If $ENC2_n \le c$, then we infer that $\hat{q}_{2m,t}$ encompasses $\hat{q}_{1m,t}$. If instead both $ENC1_n$ and $ENC2_n \geqslant c$, then the final conclusion is that $\hat{q}_{1m,t}$ does not encompass $\hat{q}_{2m,t}$ and $\hat{q}_{2m,t}$ does not encompass $\hat{q}_{1m,t}$. The conditional encompassing test for quantile forecasts can be easily generalized to the comparison of r forecasts (or, more generally, r weights). In this case, the limiting distribution of the test statistic will be χ_r^2.

One important application of our CQFE test is in the context of real-time forecast selection, that is, for selecting at time T a best forecast method for time $T+1$. To this end, we propose the following decision rule. Perform the two tests of H_{10} ($\hat{q}_{1m,t}$ encompasses $\hat{q}_{2m,t}$) and H_{20} ($\hat{q}_{2m,t}$ encompasses $\hat{q}_{1m,t}$) on data up to time T. There are four possible scenarios: (1) If neither H_{10} nor H_{20} are rejected, then the test is not helpful for forecast selection (one could, e.g., decide to use the more parsimonious model); (2) if H_{10} is rejected while H_{20} is not rejected, then one would choose $\hat{q}_{2m,T}$; (3) if H_{20} is rejected while H_{10} is not rejected, then one would choose $\hat{q}_{1m,T}$; (4) if both H_{10} and H_{20} are rejected, then one would choose the combination forecast $\hat{q}_{m,T}^c \equiv \hat{\theta}_{1n}\hat{q}_{1m,T} + \hat{\theta}_{2n}\hat{q}_{2m,T}$, where $\hat{\theta}_{1n}$ and $\hat{\theta}_{2n}$ are out-of-sample estimates of the combination weights.

4. MONTE CARLO EVIDENCE

We investigate the performance of our CQFE test in finite samples of sizes typically available to financial economists. We perform the evaluation along three dimensions: the size of the test, its power, and its sensitivity to the choice of τ in the construction of $\hat{\boldsymbol{\gamma}}_{n,\tau}$ in (17). We design our Monte Carlo experiment to match the problem of VaR evaluation and combination that is the object of our empirical application. For simplicity, we restrict attention to the conditional autoregressive value at risk (CAViaR) family of VaR models proposed by Engle and Manganelli (2004). Our choices of models within the CAViaR family and the parameter values used for the simulation are driven by the empirical application.

4.1 Size Properties

We consider forecasts generated by the asymmetric absolute value (AAV) CAViaR model,

$$VaR_{\mathrm{AAV},t+1} = \beta_0 + \beta_1 VaR_{\mathrm{AAV},t} + \beta_2|r_t - \beta_3|, \quad (18)$$

and by the symmetric absolute value (SAV) model,

$$VaR_{\mathrm{SAV},t+1} = \tilde{\beta}_0 + \tilde{\beta}_1 VaR_{\mathrm{SAV},t} + \tilde{\beta}_2|r_t|, \quad (19)$$

where $VaR_{\mathrm{AAV},t+1}$ and $VaR_{\mathrm{SAV},t+1}$ are forecasts of the conditional α-quantile of $-r_{t+1}$. Our null hypothesis is that the AAV model encompasses the SAV model. To generate data that support the null hypothesis, we proceed as follows. First, we fix the values of the true parameters $(\beta_0, \beta_1, \beta_2, \beta_3)$ and $(\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\beta}_2)$ in (18) and (19), and replicate $(VaR_{\mathrm{AAV},1}, \ldots, VaR_{\mathrm{AAV},n})$ and $(VaR_{\mathrm{SAV},1}, \ldots, VaR_{\mathrm{SAV},n})$ by assuming that $r_t \sim iid\mathcal{N}(0, \sigma^2)$ with $\sigma = .1$. In this particular case, the in-sample size m is 0 and $T = n$. Accordingly, all inference is done conditional on the set of true parameter values $(\beta_0, \beta_1, \beta_2, \beta_3)$ and $(\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\beta}_2)$. Next, we constrain $VaR_{\mathrm{AAV},t+1}$ to be the conditional α-quantile of $-r_{t+1}$ by redefining the original series. For every t, $t = 0, \ldots, n-1$, we let the data-generating process (DGP) be

$$r_{t+1} = -VaR_{\mathrm{AAV},t+1} + u_{t+1}, \quad (20)$$

with $u_{t+1} \sim iid\mathcal{N}(-\sigma\Phi^{-1}(\alpha), \sigma^2)$, $\sigma = .1$, where Φ is the distribution function of a standard normal random variable. By restricting u_{t+1} to have the α-quantile of 0, we ensure that the AAV model in (18) produces forecasts of the true conditional α-quantile of $-r_{t+1}$.

The parameter values $(\beta_0, \beta_1, \beta_2, \beta_3) = (0, .8, .3, 1)$ in (18) and $(\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\beta}_2) = (0, .9, .2)$ in (19) are chosen so as to match the estimates obtained in the empirical application for $\alpha = 5\%$.

We consider a range of values for the out-of-sample size n and the parameter τ in (17): $n = (1{,}000, 2{,}500, 5{,}000)$ and τ ranges from $.2 \times 10^{-2}$ to 10^{-2} in increments of $.1 \times 10^{-2}$. For each sample size n, we generate 30,000 Monte Carlo replications of the time series $\{r_t\}_{t=1}^n$, $\{VaR_{\mathrm{AAV},t}\}_{t=1}^n$, and $\{VaR_{\mathrm{SAV},t}\}_{t=1}^n$, each of length n. We then consider the forecast combination $(\theta_0 + \theta_{\mathrm{AAV}} \cdot VaR_{\mathrm{AAV},t} + \theta_{\mathrm{SAV}} \cdot VaR_{\mathrm{SAV},t})$ and construct the GMM estimator $(\hat{\theta}_{0n}, \hat{\theta}_{\mathrm{AAV}n}, \hat{\theta}_{\mathrm{SAV}n})'$ of the optimal weight vector $(\theta_0^*, \theta_{\mathrm{AAV}}^*, \theta_{\mathrm{SAV}}^*)'$ according to the procedure described in Section 3. Note that we include a constant term in the forecast combination, thus allowing the empirical coverage of the original forecasts to be different than the 5% nominal value. In our particular case, the AAV forecasts will display correct empirical coverage by construction, whereas the forecasts from the misspecified SAV model will in general be biased. Finally, we compute the proportion of rejections, at the 5% nominal level, of the null hypothesis

Table 1. Empirical Size of Nominal .05 Test

n	Rejection probabilities with true f_t	τ								
		.002	.003	.004	.005	.006	.007	.008	.009	.010
1,000	.040	.143	.135	.125	.116	.107	.098	.089	.082	.074
2,500	.038	.095	.085	.076	.070	.062	.054	.050	.046	.042
5,000	.037	.066	.056	.049	.042	.038	.033	.029	.025	.023

NOTE: Empirical size of the CQFE test for a nominal size of .05. Rejection frequencies are computed over 30,000 Monte Carlo replications of the null hypothesis that forecasts from the AAV CAViaR model encompass forecasts from the SAV CAViaR model when the DGP is the AAV CAViaR. n is the sample size, and τ is a user-defined constant required in the computation of our estimator of γ in (17).

$H_{10}:(\theta^*_{\mathrm{AAV}}, \theta^*_{\mathrm{SAV}}) = (1, 0)$. The test statistic $ENC1_n$ is that of Theorem 1, with $\hat{\mathbf{\Omega}}$ substituted by $\mathbf{R}\hat{\mathbf{\Omega}}\mathbf{R}'$, so as to reflect the appropriate parameter restrictions. The information vector $\mathbf{W}^*_t$ is $\mathbf{W}^*_t \equiv (1, r_t, VaR_{\mathrm{AAV},t}, VaR_{\mathrm{SAV},t})'$. The results are collected in Table 1.

The nominal 5% test appears to be well sized, with rejection probabilities around 4% across all sample sizes n, when we estimate $\boldsymbol{\gamma}$ in (14) using the true conditional density f_t of r_{t+1} in (20). In a more plausible setup in which the true density f_t is unknown and where we estimate $\boldsymbol{\gamma}$ by using our estimator $\hat{\boldsymbol{\gamma}}_{n,\tau}$ in (17), the empirical rejection probabilities vary with the sample size n and the smoothing parameter τ. A general pattern that emerges from Table 1 is that the test is oversized for $n = 1{,}000$ and small values of τ $(.2 \times 10^{-2})$ and is moderately undersized for $n = 5{,}000$ and large values of τ (10^{-2}). For other combinations of n and τ, the test appears generally well sized.

4.2 Power Properties

To generate data under the alternative hypothesis of no encompassing of AAV forecasts with respect to SAV forecasts, we first replicate $(VaR_{\mathrm{AAV},1}, \ldots, VaR_{\mathrm{AAV},n})$ and $(VaR_{\mathrm{SAV},1}, \ldots, VaR_{\mathrm{SAV},n})$ for parameter values $(\beta_0, \beta_1, \beta_2, \beta_3) = (0, .8, .3, 1)$ and $(\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\beta}_2) = (0, .9, .2)$, following the procedure described in the previous section, and then let the DGP be

$$r_{t+1} = -[\rho VaR_{\mathrm{SAV},t+1} + (1-\rho)VaR_{\mathrm{AAV},t+1}] + u_{t+1}, \quad (21)$$

where $0 < \rho < 1$ and $u_{t+1} \sim iid\mathcal{N}(-\sigma\Phi^{-1}(\alpha), \sigma^2)$, $\sigma = .1$, as in the previous section. Note that the size study is obtained when the data are generated according to (21) with $\rho = 0$. Accordingly, increasing ρ toward 1 allows us to obtain the power curve for the CQFE test. We consider a number of different values for ρ, ranging from $\rho = .02$ to $\rho = 1$, in increments of .02. For each parameterization, we generate 30,000 Monte Carlo replications of the time series $\{r_t\}_{t=1}^n$, $\{VaR_{\mathrm{AAV},t}\}_{t=1}^n$, and $\{VaR_{\mathrm{SAV},t}\}_{t=1}^n$ and proceed as previously by computing the proportion of rejections of the null hypothesis that $VaR_{\mathrm{AAV},t+1}$ encompasses $VaR_{\mathrm{SAV},t+1}$ at the 5% nominal level. Figure 1(a) plots the power curves for $n = (1{,}000, 2{,}500, 5{,}000)$ when using the true conditional density f_t in the expression (14) for $\boldsymbol{\gamma}$. As expected, the power increases with n. The loss of power induced by estimating $\boldsymbol{\gamma}$ with our estimator $\hat{\boldsymbol{\gamma}}_{n,\tau}$ in (17) is shown in Figure 1(b) for the case where $n = 2{,}500$ and for different values of the smoothing parameter τ. This figure highlights the trade-off between size and power when choosing a particular value of τ. For example, high values of τ (10^{-2}) give a well-sized test (4.2% empirical size) but with low power (40% of rejections of H_{10} when H_{20} is true), whereas low values of τ $(.2 \times 10^{-2})$ result in better power (70% of rejections of H_{10} when H_{20} is true) at the expense of size distortions (9.5% empirical size).

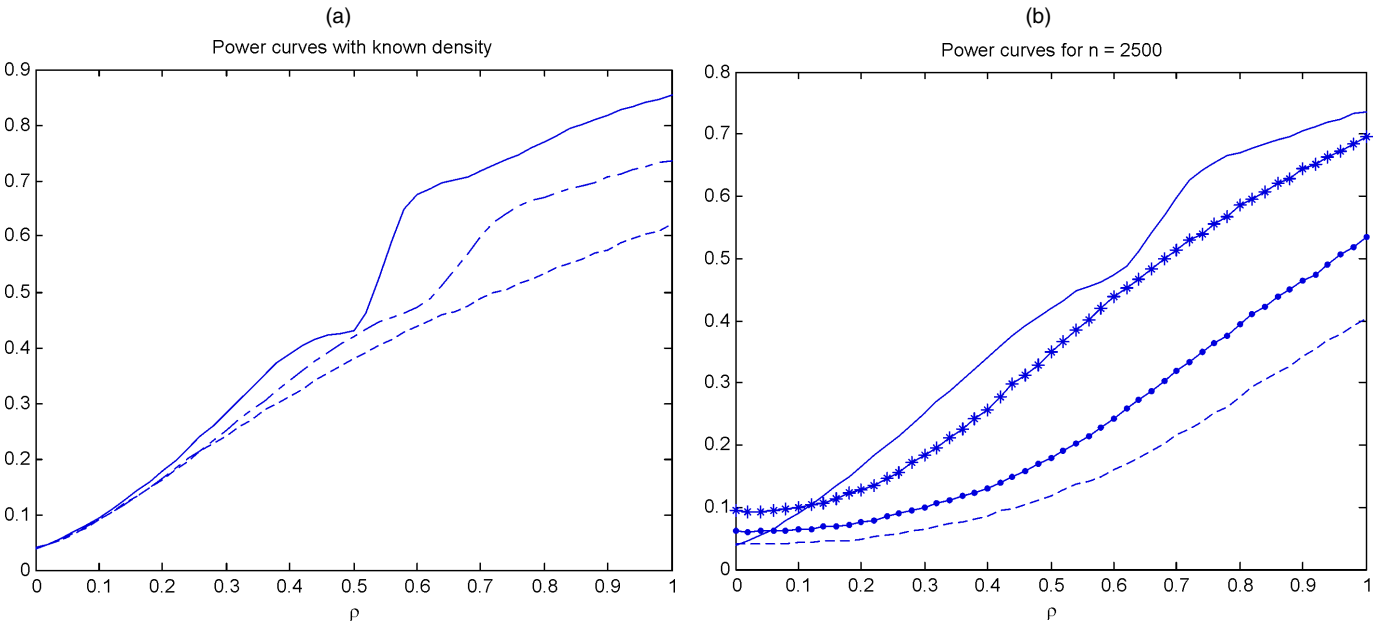

Figure 1. Power Curves of the CQFE Test in the Monte Carlo Experiment Discussed in Section 4.2 for (a) Known Density and (b) n = 2,500. Each curve represents the rejection frequency—computed by assuming f_t in (14) known—over 30,000 Monte Carlo replications. The null hypothesis being tested is that forecasts from the AAV CAViaR model encompass forecasts from the SAV CAViaR model when the DGP is a convex combination of the two, with weights ρ and $(1-\rho)$. [(a), - - - - n = 1,000; — — — n = 2,500; ——— n = 5,000; (b) ——— known density; - - - - estimated density with $\tau = .010$; ——•—— estimated density with $\tau = .006$; ———— estimated density with $\tau = .002$.]*

5. EMPIRICAL EVALUATION AND COMBINATION OF VALUE–AT–RISK FORECASTS

Here we illustrate the potential usefulness of our CQFE test by applying it to the problem of VaR evaluation. The importance of VaR became institutional in August 1996, when U.S. bank regulators adopted a "market risk" supplement to the Basle Accord of 1988. VaR has thus become a risk measure for setting capital-adequacy standards of U.S. commercial banks. The data used in our empirical application consist of 16 years of daily returns on the S&P500 index (source: Datastream), from September 1985 to September 2001 ($T = 4{,}176$ observations). The first third of the sample, corresponding to the period from September 1985 to January 1991 ($m = 1{,}392$ observations) is used as the in-sample period, while the remaining two thirds ($n = 2{,}784$ observations) are reserved to evaluate the out-of-sample performance. We adopt a fixed forecasting scheme, which means that all forecasts depend on the same set of parameters estimated over the first m observations. We consider a portfolio consisting of a long position in the index, with an investment horizon of 1 day.

5.1 VaR Models

We consider the 5% and 1% VaR forecasts originated from four different models: $VaR_{1,t+1}$ and $VaR_{2,t+1}$ are VaR forecasts based on conditional heteroscedasticity models, $r_{t+1}|\mathcal{F}_t \sim \mathcal{D}(0, \sigma^2_{t+1})$ with $\mathcal{D}$ belonging to a location-scale family of distributions. In this case VaR is a linear function of the conditional volatility of the returns σ_{t+1}, and different VaR models correspond to different specifications for the conditional variance σ^2_{t+1}. Two such specifications are the commonly used GARCH(1, 1) model, in which $\sigma^2_{1,t+1} = \omega_0 + \omega_1\sigma^2_{1,t} + \omega_2 r^2_t$, and the J. P. Morgan (1996) RiskMetrics model, where the variance is obtained as an exponential filter $\sigma^2_{2,t+1} = \lambda\sigma^2_{2,t} + (1-\lambda)r^2_t$, with $\lambda = .94$ for daily data. In both cases the corresponding VaR model is

$$VaR_{i,t+1} = \beta_0 + \beta_1\sigma_{i,t+1}, \qquad i = 1, 2. \tag{22}$$

Models such as (22) have been studied by Christoffersen et al. (2001), among others. Hereafter, we refer to $VaR_{1,t+1}$ as GARCH VaR and to $VaR_{2,t+1}$ as RiskMetrics VaR.

A different approach to VaR modeling and estimation was taken by Engle and Manganelli (2004). Here we consider two examples of the CAViaR model proposed by these authors. $VaR_{3,t+1}$ is a forecast based on an asymmetric absolute value (AAV) model,

$$VaR_{3,t+1} = \beta_0 + \beta_1 VaR_{3,t} + \beta_2|r_t - \beta_3|, \tag{23}$$

whereas $VaR_{4,t+1}$ is based on an asymmetric slope (AS) model,

$$VaR_{4,t+1} = \beta_0 + \beta_1 VaR_{4,t} + \beta_2 r^+_t + \beta_3 r^-_t, \tag{24}$$

where r^+_t and r^-_t correspond to the positive and negative parts of r_t. The three models $VaR_{1,t+1}$, $VaR_{3,t+1}$, and $VaR_{4,t+1}$ are chosen on the basis of their individual performance in modeling the VaR for the S&P500 index. As shown by Christoffersen et al. (2001), the GARCH VaR $VaR_{1,t+1}$ is the only VaR measure among several alternatives considered by the authors that passes the Christoffersen (1998) "conditional coverage test" for both 5% and 1% coverage rates. Similarly, Engle and Manganelli (2004) showed that the AAV model $VaR_{3,t+1}$ and the AS model $VaR_{4,t+1}$ are the best CAViaR specifications for the S&P500 according to a criterion that they proposed. Finally, the J. P. Morgan (1996) RiskMetrics model $VaR_{2,t+1}$ is chosen as a benchmark model commonly used by practitioners. Figures 2 and 3 show the out-of-sample sequences of VaR forecasts generated by the foregoing models, together with the sequences of VaR violations.

For each of the four VaR models (22)–(24), we first construct an estimator, $\hat{\boldsymbol{\beta}}_m \equiv \hat{\boldsymbol{\beta}}_{m,m}$, of the unknown parameter vector $\boldsymbol{\beta}$ by using the first $m = 1{,}392$ observations. We then use this estimator to form out-of-sample VaR forecasts according to a fixed forecasting scheme. In other words, at each out-of-sample date t, $m \le t \le T-1$, we compute one-step-ahead VaR forecasts, $VaR_{i,t+1}$, $i = 1, 2, 3, 4$, based on the four models (22)–(24). The computation is done recursively, meaning that for each $i = 1, 2, 3, 4$, the value of $VaR_{i,t+1}$ depends on the past forecast $VaR_{i,t}$ [$\sigma^2_{i,t}$ in the case of models (22)] and on the out-of-sample realization r_t (resp. r^2_t). For illustration, we report the parameter estimates $\hat{\boldsymbol{\beta}}_m$ in Table 2. Alternatively, one could consider sequences of VaR forecasts provided by different groups of outside researchers/analysts without knowing the underlying forecasting models, as long as the latter satisfy our assumptions.

As a quick check of the out-of-sample performance of individual VaR models and their equally weighted pairwise combinations ($.5 \cdot VaR_{i,t+1} + .5 \cdot VaR_{j,t+1}$), we compute the empirical coverage a, of the corresponding sequence of forecasts, $a \equiv n^{-1}\sum_{t=1}^{n} I_{t+1}$, where I_{t+1} denotes the "hit" variable $I_{t+1} \equiv \mathbb{1}(Y_{t+1} - VaR_{t+1} < 0)$. If the VaR model under consideration performs well, then we expect it to display correct unconditional coverage, attained when the empirical coverage a equals the nominal coverage α. Note that one could devise a simple likelihood ratio test of the null hypothesis that I_{t+1} is Bernoulli(α), which is the main principle of the so-called "unconditional coverage" test discussed by, among others, Christoffersen (1998). But this test assumes away parameter estimation uncertainty, and thus we decided not to report its results here. The out-of-sample empirical coverages are reported in Table 3.

Based on the results from Table 3, we can compare VaR models in terms of the difference between their out-of-sample empirical coverage a and the nominal coverage α. For $\alpha = 1\%$, the best model is GARCH(1, 1) with empirical coverage .853%, followed by three equally performing models with coverage .742%: AAV and equally weighted combinations of GARCH with RiskMetrics and AS. For $\alpha = 5\%$, the best empirical coverage (4.970%) is that of RiskMetrics, followed by an equally weighted combination of RiskMetrics and GARCH (4.711%) and GARCH alone (4.674%). It is interesting to note that in general, the unconditional coverage of equally weighted combinations ($.5 \cdot VaR_{i,t+1} + .5 \cdot VaR_{j,t+1}$) is between that of $VaR_{i,t+1}$ and $VaR_{j,t+1}$.

To assess the relative performance of the two models with the best empirical coverages, as identified earlier, we perform our CQFE test. Specifically, we test whether (1) at $\alpha = 1\%$ level, GARCH encompasses AAV, and (2) at $\alpha = 5\%$ level, RiskMetrics encompasses GARCH. Note that before applying the

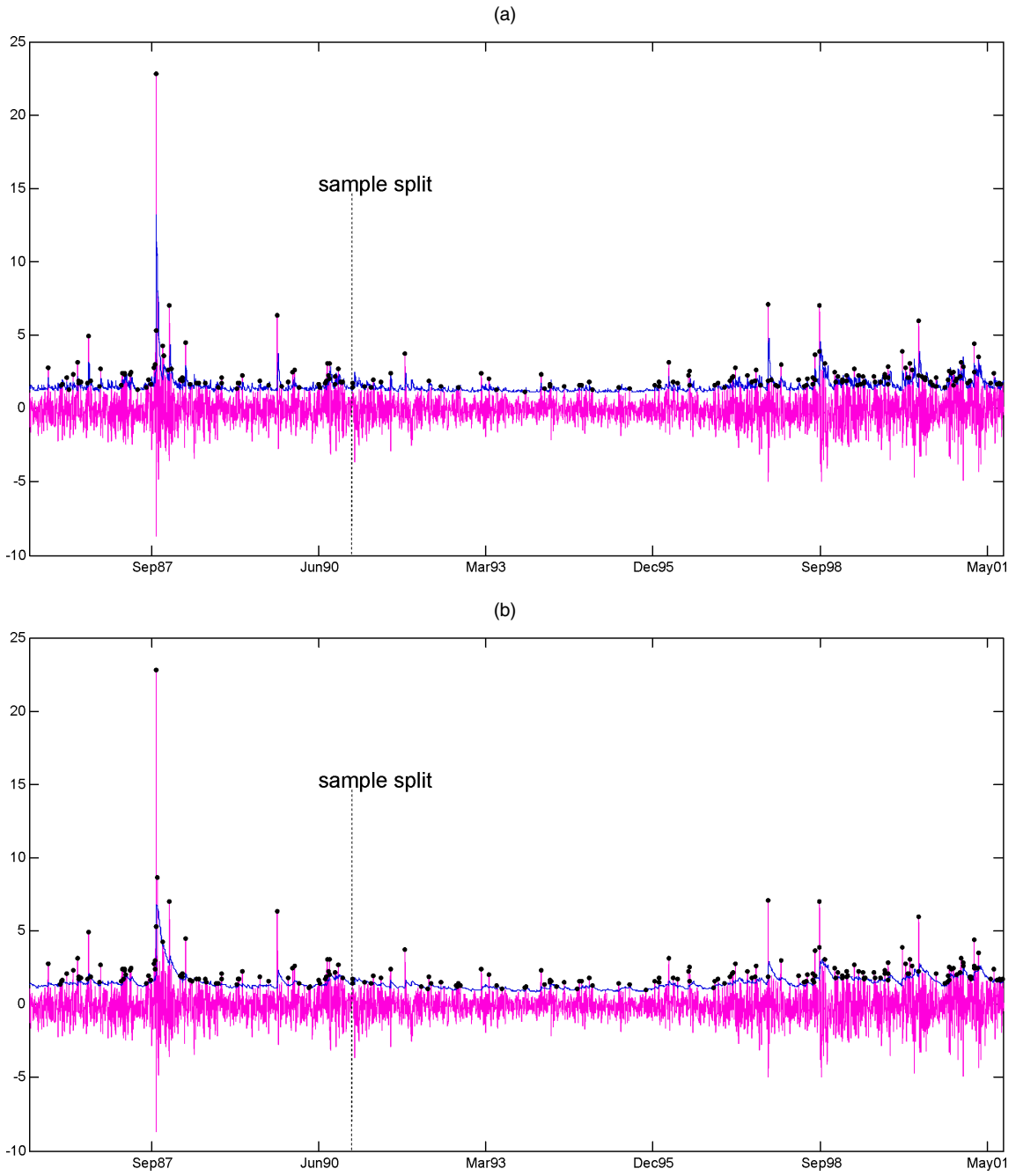

Figure 2. In- and Out-of-Sample Daily Series of Percentage Losses on S&P500 Index With 5% VaR From the (a) GARCH VaR and (b) RiskMetrics Models. VaR violations (or hits) are represented by dots.

CQFE test, we must verify that the sequences of forecasts are not perfectly correlated. The out-of-sample correlation coefficients are .91 for case (1) and .86 for case (2), which allows us to conclude that assumption (b) in Proposition 1 is not violated.

5.2 CQFE Test Results

We estimate the optimal combination weights $(\theta_0^*, \theta_i^*, \theta_j^*)'$ in the forecast combination $\theta_0 + \theta_i VaR_{i,t} + \theta_j VaR_{j,t}$ using the GMM approach described in Section 3. For the purposes of this empirical application, we let $\mathbf{W}_t^* \equiv (1, r_t, VaR_{i,t}, VaR_{j,t})'$.

We report the estimated combination weights $\hat{\theta}_{0n}$, $\hat{\theta}_{in}$, and $\hat{\theta}_{jn}$ together with their standard errors, in Table 4. It is important to note that the computation of standard errors is based on our estimator $\hat{\boldsymbol{\gamma}}_{n,\tau}$ given in (17), in which the smoothing parameter τ takes values $.2 \times 10^{-2}$, $.6 \times 10^{-2}$, and 10^{-2}. For these values of τ, the CQFE test has reasonable size and power properties, as shown in the Monte Carlo exercise. Table 4 also contains the corresponding values of the test statistics ENC_{in} and ENC_{jn}.

As can be seen from Table 4, neither forecast encompasses its competitor for both levels of α. This implies that the forecast combination in both cases outperforms the individual forecasts. However, note that for $\alpha = 5\%$, the weight on the RiskMetrics forecast is not significantly different from 0 (t-statistics range from .046 to .074), suggesting that the optimal combination in this case is simply the bias-corrected GARCH forecast.

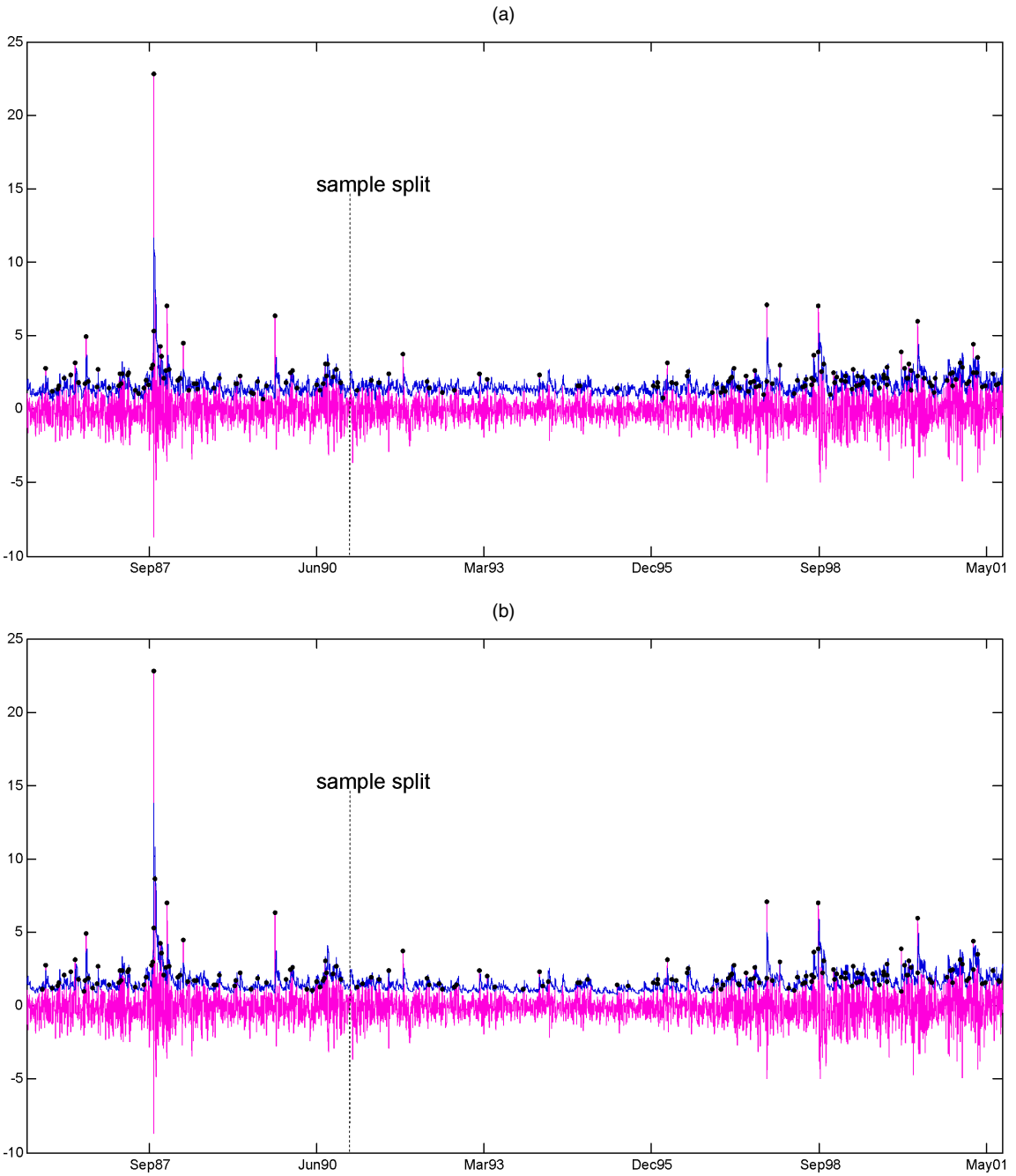

Figure 3. In- and Out-of-Sample Daily Series of Percentage Losses on S&P500 Index With 5% VaR From the (a) AAV and (b) AS Models. VaR violations (or hits) are represented by dots.

6. CONCLUSION

In this article we propose a CQFE test for comparing alternative conditional quantile forecasts in an out-of-sample framework. We base our evaluation on the concept of encompassing, which requires that a forecast be able to explain the predictive ability of a rival forecast. The CQFE test thus can be viewed as a test of superior predictive ability. The setup proposed in this article also allows us to discuss the benefit of forecast combination for quantile forecasts, which becomes relevant when the encompassing tests indicate that neither forecast outperforms its competitor.

The key features of our approach are (1) the use of the tick loss function rather than the quadratic loss function in the definition of encompassing; (2) a conditional, rather than unconditional, approach to out-of-sample evaluation; and (3) the derivation of our test in an environment with asymptotically nonvanishing estimation uncertainty. Some of the benefits of our approach are that it allows comparison of forecasts based on both nested and nonnested models and of forecasts produced by general estimation procedures.

Implementation of the CQFE test is done using a fairly standard GMM estimation technique, with the optimization procedure appropriately modified to accommodate our nondif-

Table 2. VaR Parameter Estimates

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
$\alpha = .01$				
GARCH	.982	1.597		
	(.048)	(.033)		
RiskMetrics	.959	1.698		
	(.104)	(.125)		
AAV	.213	.714	.761	.422
	(.020)	(.040)	(.068)	(.026)
AS	.460	.716	.110	−.796
	(.029)	(.061)	(.011)	(.081)
$\alpha = .05$				
GARCH	.055	1.446		
	(.075)	(.095)		
RiskMetrics	.500	1.039		
	(.104)	(.137)		
AAV	−.074	.804	.328	1.070
	(.008)	(.016)	(.015)	(.065)
AS	.120	.834	.025	−.404
	(.048)	(.058)	(.006)	(.123)

NOTE: Parameter estimates for different VaR models. Data are Datastream daily returns on S&P500 from September 1985 to January 1991 ($m = 1{,}392$ observations). The estimation is carried out by GMM in the GARCH and RiskMetrics VaR models and by QML in the CAViaR models. For VaR models where $r_{t+1}|F_t \sim \mathcal{N}(0, \sigma^2_{t+1})$ with (1) GARCH volatility $\sigma^2_{t+1} = \omega_0 + \omega_1\sigma^2_t + \omega_2 r^2_t$ we have $\omega_0 = .117$, $\omega_1 = .763$, and $\omega_2 = .150$, and for those with (2) RiskMetrics volatility $\sigma^2_{t+1} = \lambda\sigma^2_t + (1-\lambda)r^2_t$ we take $\lambda = .94$.

ferentiable criterion function. The CQFE test displays good size and power properties for samples of sizes typically available in financial applications.

We apply the CQFE test to the problem of conditional VaR forecast evaluation using S&P500 daily index returns. At the 1% level, we find that a forecast combination (with intercept) of GARCH and AAV CAViaR forecasts outperforms both individual components. A similar result holds at 5% level, where we compare VaR forecasts generated from RiskMetrics and GARCH models. In the latter case, however, we find that the combination weight on the RiskMetrics forecast is not significantly different from 0, indicating that bias-corrected GARCH forecasts for the 5% VaR encompass RiskMetrics forecasts.

ACKNOWLEDGMENTS

The authors thank Graham Elliott, Clive Granger, Jose Lopez, Andrew Patton, and Kevin Sheppard, as well as the participants to the 2003 ASSA meeting in Washington, DC, UTS workshop, and Duke Conference on Forecasting, for their valuable comments and suggestions, and Peter Christoffersen and Eric Ghysels for providing their data. They also thank the editors, the associate editor, and two anonymous referees for their useful comments that led to a considerably improved version of the article. Any remaining errors are the authors' own.

Table 3. Out-of-Sample Empirical Coverage

	GARCH	RiskMetrics	AAV	AS
$\alpha = .01$				
GARCH	.853%	.742%	.705%	.742%
RiskMetrics		.705%	.705%	.631%
AAV			.742%	.668%
AS				.631%
$\alpha = .05$				
GARCH	4.674%	4.711%	4.191%	4.191%
RiskMetrics		4.970%	4.228%	4.191%
AAV			4.303%	4.303%
AS				4.228%

NOTE: Empirical coverage $a = n^{-1}\sum I_{t+1}$ for individual VaR models (diagonal elements) and their equally weighted pairwise combinations (off-diagonal elements). Data: Datastream daily returns on S&P500 from January 1991 to September 2001 ($n = 2{,}784$ observations).

APPENDIX: PROOFS

We use the following notation throughout. If $\mathbf{V}$ is a real n-vector, $\mathbf{V} \equiv (V_1, \ldots, V_n)'$, then $\|\mathbf{V}\|$ denotes the L_2-norm of $\mathbf{V}$, that is, $\|\mathbf{V}\|^2 \equiv \mathbf{V}'\mathbf{V} = \sum_{i=1}^n V_i^2$. If $\mathbf{M}$ is a real $n \times n$-matrix, $\mathbf{M} \equiv (M_{ij})_{1\le i,j\le n}$, then $\|\mathbf{M}\|$ denotes the L_∞-norm of $\mathbf{M}$, that is, $\|\mathbf{M}\| \equiv \max_{1\le i,j\le n}|M_{ij}|$.

Proof of Lemma 1

Let

$$\begin{aligned}\Sigma_t(\boldsymbol{\theta}) &\equiv E_t\big[\big(\alpha - \mathbb{1}(Y_{t+1} - \boldsymbol{\theta}'\hat{\mathbf{q}}_{m,t} < 0)\big)(Y_{t+1} - \boldsymbol{\theta}'\hat{\mathbf{q}}_{m,t})\big] \\ &= \int_{\mathbb{R}} \alpha(y_{t+1} - \boldsymbol{\theta}'\hat{\mathbf{q}}_{m,t})\, dF_t(y_{t+1}) \\ &\quad - \int_{\mathbb{R}} \mathbb{1}(y_{t+1} - \boldsymbol{\theta}'\hat{\mathbf{q}}_{m,t} < 0)(y_{t+1} - \boldsymbol{\theta}'\hat{\mathbf{q}}_{m,t})\, dF_t(y_{t+1}) \\ &= \int_{-\infty}^{+\infty} \alpha(y_{t+1} - \boldsymbol{\theta}'\hat{\mathbf{q}}_{m,t})\, dF_t(y_{t+1}) \\ &\quad - \int_{-\infty}^{0} x_{t+1}\, dF_t(x_{t+1} + \boldsymbol{\theta}'\hat{\mathbf{q}}_{m,t}),\end{aligned}$$

Table 4. Conditional Quantile Forecast Encompassing Test for VaR Measures

Model	$\hat{\theta}_{0n}$	$\hat{\theta}_{1n}$	$\hat{\theta}_{2n}$	J	ENC_{1n}	ENC_{2n}
$\alpha = .01$						
GARCH(1) versus AAV(2)	−1.506	1.048	.382	8.636		
$\tau = .002$	(.248)	(.147)	(.748)		70.779*	69.224*
$\tau = .006$	(.298)	(.158)	(.074)		53.104*	71.181*
$\tau = .010$	(.392)	(.209)	(.089)		40.071*	48.335*
$\alpha = .05$						
RiskMetrics(1) versus GARCH(2)	−.118	.005	.565	10.545		
$\tau = .002$	(.057)	(.068)	(.092)		602.510*	150.690*
$\tau = .006$	(.066)	(.086)	(.102)		260.544*	96.291*
$\tau = .010$	(.083)	(.109)	(.125)		152.563*	61.207*

NOTE: Out-of-sample CQFE test for VaR measures for a portfolio composed of a long position in S&P500 index with an investment horizon of 1 day. Data are Datastream daily returns on S&P500 from January 1991 to September 2001 ($n = 2{,}784$ observations). The consistent standard errors of the GMM estimator $(\theta_{0n}, \theta_{1n}, \theta_{2n})'$ were computed with $\tau = .002, .006, .010$ and are reported in parentheses. J is the value of the J-test statistics: $J = \mathbf{g}_n(\boldsymbol{\theta}_n)'\mathbf{S}^{-1}\mathbf{g}_n(\boldsymbol{\theta}_n)$. The marked (*) values of the CQFE test statistics ENC_{1n} and ENC_{2n} are significant at the 1% level.

where $x_{t+1} \equiv y_{t+1} - \boldsymbol{\theta}'\hat{\mathbf{q}}_{m,t}$. Thus $\nabla_{\boldsymbol{\theta}} \Sigma_t(\boldsymbol{\theta}) = -\alpha\hat{\mathbf{q}}_{m,t} - \int_{-\infty}^{0} \hat{\mathbf{q}}_{m,t} x_{t+1} f_t(x_{t+1} + \boldsymbol{\theta}'\hat{\mathbf{q}}_{m,t})\, dx_{t+1}$, because we assume that the random variable Y_{t+1} has a continuously differentiable conditional density f_t, that is, $dF_t(y_{t+1}) = f_t(y_{t+1})\, dy_{t+1}$ and f_t continuous. By arranging the previous equality, we obtain $\nabla_{\boldsymbol{\theta}} \Sigma_t(\boldsymbol{\theta}) = -\alpha\hat{\mathbf{q}}_{m,t} - [\hat{\mathbf{q}}_{m,t} x_{t+1} f_t(x_{t+1} + \boldsymbol{\theta}'\hat{\mathbf{q}}_{m,t})]_{-\infty}^{0} + \int_{-\infty}^{0} \hat{\mathbf{q}}_{m,t} f_t(x_{t+1} + \boldsymbol{\theta}'\hat{\mathbf{q}}_{m,t})\, dx_{t+1}$, so that $\nabla_{\boldsymbol{\theta}} \Sigma_t(\boldsymbol{\theta}) = -\alpha\hat{\mathbf{q}}_{m,t} + \hat{\mathbf{q}}_{m,t} \int_{-\infty}^{\boldsymbol{\theta}'\hat{\mathbf{q}}_{m,t}} f_t(y_{t+1})\, dy_{t+1}$. We can then write $\nabla_{\boldsymbol{\theta}} \Sigma_t(\boldsymbol{\theta}) = -E_t[(\alpha - \mathbb{1}(Y_{t+1} - \boldsymbol{\theta}'\hat{\mathbf{q}}_{m,t} < 0))\hat{\mathbf{q}}_{m,t}]$. If $\boldsymbol{\theta}_t^*$ is a solution to the initial minimization problem, then $\nabla_{\boldsymbol{\theta}} \Sigma_t(\boldsymbol{\theta})|_{\boldsymbol{\theta}_t^*} = 0$, a.s.-$P$, that is,

$$E_t\big[\big(\alpha - \mathbb{1}(Y_{t+1} - \boldsymbol{\theta}_t^{*\prime}\hat{\mathbf{q}}_{m,t} < 0)\big)\hat{\mathbf{q}}_{m,t}\big] = 0, \quad \text{a.s.-}P.$$

Because $\hat{q}_{m,t}$ is $\mathcal{F}_t$-measurable, we can rewrite the previous equation as $E_t[\alpha - \mathbb{1}(Y_{t+1} - \boldsymbol{\theta}_t^{*\prime}\hat{\mathbf{q}}_{m,t} < 0)] = 0$, a.s.-$P$, CQFD.

Lemma A.1. For all t, $m \le t \le T-1$, if $\text{corr}(\hat{q}_{1m,t}, \hat{q}_{2m,t}) \neq \pm 1$ and $\hat{q}_{im,t} \neq 0$, a.s.-P for $i = 1, 2$ then $\hat{q}_{1m,t}$ and $\hat{q}_{2m,t}$ are linearly independent; that is, $\gamma_1\hat{q}_{1m,t} + \gamma_2\hat{q}_{2m,t} = 0$, a.s.-$P$ implies $\gamma_1 = \gamma_2 = 0$.

Proof of Lemma A.1. By contradiction, suppose there exist $(\gamma_1, \gamma_2) \neq (0, 0)$ such that $\gamma_1\hat{q}_{1m,t} + \gamma_2\hat{q}_{2m,t} = 0$, a.s.-$P$. Without loss of generality, suppose that $\gamma_1 \neq 0$. Then $\hat{q}_{1m,t} = -(\gamma_2/\gamma_1)\hat{q}_{2m,t}$, a.s.-$P$, from which it follows that either (1) $\gamma_2 = 0$, which implies that $\hat{q}_{1m,t} = 0$, a.s.-P, or (2) $\gamma_2 \neq 0$, which implies that $\text{corr}(\hat{q}_{1m,t}, \hat{q}_{2m,t}) = \text{sgn}(-(\gamma_2/\gamma_1)) = \pm 1$, CQFD.

Lemma A.2. Under assumptions (a) and (b) of Proposition 1, θ^* is unique.

Proof of Lemma A.2. We show that $\mathbf{0} = E[\mathbf{g}(\boldsymbol{\theta}^*; Y_{t+1}, \mathbf{W}_t^*)] = E[\mathbf{g}(\bar{\boldsymbol{\theta}}; Y_{t+1}, \mathbf{W}_t^*)] \Rightarrow \boldsymbol{\theta}^* = \bar{\boldsymbol{\theta}}$. Consider the difference $\Delta(\mathbf{W}_t^*)$, defined by $\Delta(\mathbf{W}_t^*) \equiv E[(\alpha - \mathbb{1}(Y_{t+1} - \boldsymbol{\theta}^{*\prime} \times \hat{\mathbf{q}}_{m,t} < 0))\mathbf{W}_t^*] - E[(\alpha - \mathbb{1}(Y_{t+1} - \bar{\boldsymbol{\theta}}'\hat{\mathbf{q}}_{m,t} < 0))\mathbf{W}_t^*]$. We have

$$\begin{aligned}\Delta(\mathbf{W}_t^*) &= E\big[\mathbf{W}_t^*\big(\mathbb{1}(Y_{t+1} - \bar{\boldsymbol{\theta}}'\hat{\mathbf{q}}_{m,t} < 0) \\ &\qquad - \mathbb{1}(Y_{t+1} - \boldsymbol{\theta}^{*\prime}\hat{\mathbf{q}}_{m,t} < 0)\big)\big] \\ &= E\big[\mathbf{W}_t^* E_t\big[\mathbb{1}(\boldsymbol{\theta}^{*\prime}\hat{\mathbf{q}}_{m,t} < Y_{t+1} < \bar{\boldsymbol{\theta}}'\hat{\mathbf{q}}_{m,t}) \\ &\qquad - \mathbb{1}(\bar{\boldsymbol{\theta}}'\hat{\mathbf{q}}_{m,t} < Y_{t+1} < \boldsymbol{\theta}^{*\prime}\hat{\mathbf{q}}_{m,t})\big]\big],\end{aligned}$$

because $\mathbf{W}_t^*$ is $\mathcal{F}_t$-measurable. The conditional expectation on the right side of the equality is in turn equal to $\int_{\boldsymbol{\theta}^{*\prime}\hat{\mathbf{q}}_{m,t}}^{\bar{\boldsymbol{\theta}}'\hat{\mathbf{q}}_{m,t}} f_t(y_{t+1})\, dy_{t+1} \equiv D_t(\bar{\boldsymbol{\theta}}, \boldsymbol{\theta}^*)$. Thus $\Delta(\mathbf{W}_t^*) = E[\mathbf{W}_t^* D_t(\bar{\boldsymbol{\theta}}, \boldsymbol{\theta}^*)]$, and we have $\Delta(\mathbf{W}_t^*) = 0 \Rightarrow D_t(\bar{\boldsymbol{\theta}}, \boldsymbol{\theta}^*) = 0$, a.s.-$P$, given that $\mathbf{W}_t^*$ incorporates all available information in $\mathcal{F}_t$. By assumption (a), $f_t(\cdot)$ is continuous and strictly positive on $\mathbb{R}$, so that $D_t(\bar{\boldsymbol{\theta}}, \boldsymbol{\theta}^*)$ can be 0 a.s.-P only when $\bar{\boldsymbol{\theta}}'\hat{\mathbf{q}}_{m,t} = \boldsymbol{\theta}^{*\prime}\hat{\mathbf{q}}_{m,t}$, a.s.-$P$, that is, $(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)'\hat{\mathbf{q}}_{m,t} = 0$, a.s.-$P$. From Lemma A.1, this implies that $(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) = \mathbf{0}$. Hence $\Delta(\mathbf{W}_t^*) = 0 \Rightarrow \boldsymbol{\theta}^* = \bar{\boldsymbol{\theta}}$, CQFD.

Proof of Proposition 1

We first discuss the nature of the sequence $\{\mathbf{g}(\boldsymbol{\theta}; Y_{t+1}, \mathbf{W}_t^*)\}$, where $\mathbf{g}(\boldsymbol{\theta}; Y_{t+1}, \mathbf{W}_t^*)$ depends on the data through $Y_{t+1}, \mathbf{W}_t^*$, and $\hat{\mathbf{q}}_{m,t}$. Consider the cases of fixed forecasting scheme and rolling window forecasting scheme separately.

For a fixed forecasting scheme, the forecasts $\hat{\mathbf{q}}_{m,t}$, $t = m, \ldots, T-1$, depend on the one hand on predetermined parameter estimates $\hat{\boldsymbol{\beta}}_{m,m}$, hence on the variables $(X_1, \ldots, X_m)$, and on the other hand on some set of right-side variables of the forecasting model that are observed at time t. Typically, those variables are included in the vector $\mathbf{W}_t^*$. Therefore, by letting $\mathbf{V}_{t+1} \equiv (Y_{t+1}, \mathbf{W}_t^*, X_1, \ldots, X_m)'$, we can rewrite $\mathbf{g}(\boldsymbol{\theta}; Y_{t+1}, \mathbf{W}_t^*)$ as $\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})$.

For a rolling window forecasting scheme, $\hat{\mathbf{q}}_{m,t}$, $t = m, \ldots, T-1$, is a constant measurable function of the estimation window, which consists of the m most recent observations of X_t. In that case, we can again let $\mathbf{V}_{t+1} \equiv (Y_{t+1}, \mathbf{W}_t^*, X_t, \ldots, X_{t-m+1})'$ and rewrite $\mathbf{g}(\boldsymbol{\theta}; Y_{t+1}, \mathbf{W}_t^*)$ as $\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})$.

Because for every t, $t = m, \ldots, T-1$, $\mathbf{V}_{t+1}$ is a function of a finite number $(m+2)$ of variables which, by assumption (c), are strictly stationary and α-mixing, the sequence $\{\mathbf{V}_t\}$ is strictly stationary and α-mixing of the same size (see, e.g., thm. 3.49 in White 2001). Note that strict stationarity and α-mixing of $\{\mathbf{V}_t\}$ imply ergodicity (see, e.g., thm. 3.44 in White 2001), so that we can use one of the standard results on the consistency of GMM estimators for stationary and ergodic sequences. Specifically, we verify that the conditions of thm. 2.6 of Newey and McFadden (1994, pp. 2132–2133) are satisfied in our case. (Note that the results of thm. 2.6 hold if the iid assumption is replaced with the condition that $\{\mathbf{V}_t\}$ is strictly stationary and ergodic.) First, we need to show that $\hat{\mathbf{S}}_n(\hat{\boldsymbol{\theta}}_n) \xrightarrow{p} \mathbf{S}$, where $\mathbf{S}$ is defined in (12) and, from (13), $\hat{\mathbf{S}}_n(\boldsymbol{\theta}) \equiv n^{-1} \sum_{t=m}^{T-1} \mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})'$. Note that $\mathbf{g}$ is an $\mathcal{F}_{t+1}$-measurable function of $\{\mathbf{V}_{t+1}\}$, which is strictly stationary and α-mixing. Using, once again, theorem 3.49 of White (2001), we then have that $\{\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})\}$ and $\{\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})'\}$ are strictly stationary and α-mixing of same size. Hence we can apply a law of large numbers (LLN) for α-mixing sequences to show that for every $\boldsymbol{\theta} \in \Theta$, $\hat{\mathbf{S}}_n(\boldsymbol{\theta})$ converges to $\tilde{\mathbf{S}}(\boldsymbol{\theta}) \equiv E[\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})']$. Specifically, we check that the assumptions of corollary 3.48 of White (2001) hold. First, note that for $r > 2$, we have $-r/(r-1) > -r/(r-2)$, so that the sequence $\{\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})'\}$ is α-mixing with α of size $-r/(r-1)$. We now need to show that for some $\tilde{\delta} > 0$, we have $E\|\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})'\|^{r+\tilde{\delta}} < \infty$. Recall from (9) that

$$\begin{aligned}&\|\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})'\| \\ &\quad = [\alpha - \mathbb{1}(Y_{t+1} - \boldsymbol{\theta}'\hat{\mathbf{q}}_{m,t} < 0)]^2 \|\mathbf{W}_t^*\mathbf{W}_t^{*\prime}\| \\ &\quad \le \|\mathbf{W}_t^*\mathbf{W}_t^{*\prime}\|, \quad \text{a.s.-}P.\end{aligned}$$

Moreover, we know (by norm equivalence) that there exist some positive constant c such that

$$\begin{aligned}\|\mathbf{W}_t^*\mathbf{W}_t^{*\prime}\| &= \big|W_{t,i_0}^* \cdot W_{t,j_0}^*\big| \\ &\le \big|W_{t,i_0}^*\big| \cdot \big|W_{t,j_0}^*\big| \le c^2 \cdot \|\mathbf{W}_t^*\|^2, \quad \text{a.s.-}P,\end{aligned}$$

where i_0 and j_0, $1 \le i_0, j_0 \le h = \dim(\mathbf{W}_t^*)$, are such that $\|\mathbf{W}_t^*\mathbf{W}_t^{*\prime}\| = \max_{1 \le i,j \le h} |W_{t,i}^* \cdot W_{t,j}^*| = |W_{t,i_0}^* \cdot W_{t,j_0}^*|$. Hence $E\|\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})'\|^{r+\tilde{\delta}} \le c^2 \cdot \max\{1, E\|\mathbf{W}_t^*\|^{2r+2\tilde{\delta}}\}$, and

so by letting $2\tilde{\delta} = \delta$ and using assumption (e), we get $E\|\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})'\|^{r+\tilde{\delta}} < \infty$. Together, the strict stationarity of $\{\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})'\}$ and corollary 3.48 of White (2001) then ensure that $\hat{\mathbf{S}}_n(\boldsymbol{\theta}) \xrightarrow{p} \tilde{\mathbf{S}}(\boldsymbol{\theta}) = E[\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})']$. In particular, if $\hat{\boldsymbol{\theta}}_n$ is some previously obtained consistent estimate of $\boldsymbol{\theta}^*$, then $\hat{\mathbf{S}}_n(\hat{\boldsymbol{\theta}}_n) \xrightarrow{p} \tilde{\mathbf{S}}(\boldsymbol{\theta}^*) = E[\mathbf{g}(\boldsymbol{\theta}^*; \mathbf{V}_{t+1})\mathbf{g}(\boldsymbol{\theta}^*; \mathbf{V}_{t+1})']$, which, due to the fact that $\{\mathbf{g}(\boldsymbol{\theta}^*; \mathbf{V}_{t+1}), \mathcal{F}_t\}$ is a martingale difference sequence, equals the asymptotic covariance matrix $\mathbf{S}$ in (12).

We now check that all the other conditions of theorem 2.6 of Newey and McFadden (1994) are satisfied; in particular, we have $\mathbf{S} = E[\mathbf{g}(\boldsymbol{\theta}^*; Y_{t+1}, \mathbf{W}_t^*)\mathbf{g}(\boldsymbol{\theta}^*; Y_{t+1}, \mathbf{W}_t^*)'] = E\{[\alpha - \mathbb{1}(Y_{t+1} - \boldsymbol{\theta}^{*\prime}\hat{\mathbf{q}}_{m,t} < 0)]^2\mathbf{W}_t^*\mathbf{W}_t^{*\prime}\}$, so that for any $\boldsymbol{\zeta} \in \mathbb{R}^h$, we have $\boldsymbol{\zeta}' \cdot \mathbf{S} \cdot \boldsymbol{\zeta} = 0$ if and only if

$$\begin{aligned}
&\boldsymbol{\zeta}'[\alpha - \mathbb{1}(Y_{t+1} - \boldsymbol{\theta}^{*\prime}\hat{\mathbf{q}}_{m,t} < 0)]^2\mathbf{W}_t^*\mathbf{W}_t^{*\prime}\boldsymbol{\zeta} \\
&\quad = [\alpha - \mathbb{1}(Y_{t+1} - \boldsymbol{\theta}^{*\prime}\hat{\mathbf{q}}_{m,t} < 0)]^2[\mathbf{W}_t^{*\prime}\boldsymbol{\zeta}]^2 \\
&\quad = 0, \quad \text{a.s.-}P,
\end{aligned}$$

which is equivalent to $\mathbf{W}_t^{*\prime}\boldsymbol{\zeta} = 0$, a.s.-$P$. Because we know from assumption (d) that $E[\mathbf{W}_t^*\mathbf{W}_t^{*\prime}]$ is nonsingular, this last equality implies that $\boldsymbol{\zeta}$ needs to be equal to an h-vector of 0's. Hence the matrices $\mathbf{S}$ and its inverse $\mathbf{S}^{-1}$ are positive definite. In particular, this implies that $\mathbf{S}^{-1}E[\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})] = \mathbf{0}$ only if $E[\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})] = \mathbf{0}$. This, together with Lemma A.2, verifies condition 2.6(i). Condition 2.6(ii) is the standard compactness condition on the parameter space Θ that we impose here. The continuity condition 2.6(iii) holds because $\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})$ is a.s. continuous on Θ. Indeed, note that the only discontinuity point occurs when $Y_{t+1} = \boldsymbol{\theta}^{*\prime}\hat{\mathbf{q}}_{m,t}$, a.s.-$P$, which, due to the continuity of Y_{t+1}, occurs with probability 0. Finally, condition 2.6(iv) is verified by imposing assumption (e), because for all $\boldsymbol{\theta} \in \Theta$, we have $\|\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})\| \le \|\mathbf{W}_t^*\|$, a.s.-$P$, so that $E[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{g}(\boldsymbol{\theta}; \mathbf{V}_{t+1})\|] \le E\|\mathbf{W}_t^*\| < \max\{1, E\|\mathbf{W}_t^*\|^{2r+\delta}\} < \infty$. Theorem 2.6 of Newey and McFadden (1994) then ensures that $\hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}^*$, CQFD.

Lemma A.3. Let the assumptions of Proposition 1 hold. We then have $\sqrt{n}\|\mathbf{g}_n(\hat{\boldsymbol{\theta}}_n)\| \xrightarrow{p} 0$.

Proof of Lemma A.3. Recall from (11) that $\hat{\boldsymbol{\theta}}_n$ minimizes $[\mathbf{g}_n(\boldsymbol{\theta})'\hat{\mathbf{S}}_n^{-1}[\mathbf{g}_n(\boldsymbol{\theta})]$ on Θ, where $\mathbf{g}_n(\boldsymbol{\theta}) = n^{-1}\sum_{t=m}^{T-1}(\alpha - \mathbb{1}(y_{t+1} - \boldsymbol{\theta}'\hat{\mathbf{q}}_{m,t} < 0))\mathbf{w}_t^*$ and $\hat{\mathbf{S}}_n$ is positive definite. Then $\hat{\boldsymbol{\theta}}_n$ also minimizes $\|\mathbf{g}_n(\boldsymbol{\theta})\|^2 = [\mathbf{g}_n(\boldsymbol{\theta})'[\mathbf{g}_n(\boldsymbol{\theta})]$. For $i = 1, 2$ and $j = 1, \ldots, h = \dim(\mathbf{W}_t^*)$, let $\hat{g}_{n,i,j}(\varepsilon) \equiv n^{-1}\sum_{t=m}^{T-1}(\alpha - \mathbb{1}(y_{t+1} - (\hat{\boldsymbol{\theta}}_n + \varepsilon\mathbf{e}_i)'\hat{\mathbf{q}}_{m,t} < 0))w_{t,j}^*$, where $\{\mathbf{e}_1, \mathbf{e}_2\}$ is the standard basis of $\mathbb{R}^2$ and $\varepsilon \in \mathbb{R}$ is such that for $i = 1, 2$, $\hat{\boldsymbol{\theta}} + \varepsilon\mathbf{e}_i \in \Theta$. Note that $\hat{g}_{n,i,j}(0) = g_{n,j}(\hat{\boldsymbol{\theta}}_n)$, where $g_{n,j}$ is the jth component of $\mathbf{g}_n$. For $i = 1, 2$ and $j = 1, \ldots, h$, the function $\varepsilon \mapsto [\hat{g}_{n,i,j}(\varepsilon)]^2$ is convex, so that for every $\varepsilon > 0$, we have

$$\begin{aligned}
[\hat{g}_{n,i,j}(0)]^2 - [\hat{g}_{n,i,j}(-\varepsilon)]^2 &\le \{[\hat{g}_{n,i,j}(\varepsilon)]^2 - [\hat{g}_{n,i,j}(-\varepsilon)]^2\}/2 \\
&\le [\hat{g}_{n,i,j}(\varepsilon)]^2 - [\hat{g}_{n,i,j}(0)]^2. \qquad \text{(A.1)}
\end{aligned}$$

Now note that

$$\begin{aligned}
&[\hat{g}_{n,i,j}(\varepsilon)]^2 - [\hat{g}_{n,i,j}(-\varepsilon)]^2 \\
&\quad = [\hat{g}_{n,i,j}(\varepsilon) + \hat{g}_{n,i,j}(-\varepsilon)] \\
&\qquad \times \left[n^{-1}\sum_{t=m}^{T-1}\big(\mathbb{1}\big(y_{t+1} - (\hat{\boldsymbol{\theta}}_n - \varepsilon\mathbf{e}_i)'\hat{\mathbf{q}}_{m,t} < 0\big) \right. \\
&\qquad\qquad \left. - \mathbb{1}\big(y_{t+1} - (\hat{\boldsymbol{\theta}}_n + \varepsilon\mathbf{e}_i)'\hat{\mathbf{q}}_{m,t} < 0\big)\big)w_{t,j}^* \right],
\end{aligned}$$

so that when $\varepsilon \to 0$, $[\hat{g}_{n,i,j}(\varepsilon)]^2 - [\hat{g}_{n,i,j}(-\varepsilon)]^2 \to 2\hat{g}_{n,i,j}(0) \times [n^{-1}\sum_{t=m}^{T-1}\mathbb{1}(y_{t+1} = \hat{\boldsymbol{\theta}}_n'\hat{\mathbf{q}}_{m,t})w_{t,j}^*]$. Using inequality (A.1), it must therefore be the case that

$$P\left(\hat{g}_{n,i,j}(0)\left[n^{-1}\sum_{t=m}^{T-1}\mathbb{1}(Y_{t+1} = \hat{\boldsymbol{\theta}}_n'\hat{\mathbf{q}}_{m,t})W_{t,j}^*\right] = 0\right) = 1. \quad \text{(A.2)}$$

Hence

$$\begin{aligned}
&P\big(\sqrt{n}\|g_n(\hat{\boldsymbol{\theta}}_n)\| > \epsilon\big) \\
&\quad \le P\Big(\max_{1\le j\le h}|g_{n,i,j}(\hat{\boldsymbol{\theta}}_n)| > \epsilon/\sqrt{n}\Big) \\
&\quad = P\Big(\max_{1\le j\le h}|\hat{g}_{n,i,j}(0)| > \epsilon/\sqrt{n}\Big) \\
&\quad \le P\left(\max_{1\le j\le h}|\hat{g}_{n,i,j}(0)| \right. \\
&\qquad \left. \times \left[n^{-1}\sum_{t=m}^{T-1}\mathbb{1}(Y_{t+1} = \hat{\boldsymbol{\theta}}_n'\hat{\mathbf{q}}_{m,t})W_{t,j}^*\right] > \epsilon/\sqrt{n}\right),
\end{aligned}$$

where we have used the fact that Y_{t+1} is a continuous random variable, so that $n^{-1}\sum_{t=m}^{T-1}\mathbb{1}(Y_{t+1} = \hat{\boldsymbol{\theta}}_n'\hat{\mathbf{q}}_{m,t})W_{t,j}^* = o_p(1)$. Using condition (A.2), finally the foregoing inequality implies that $\sqrt{n}\|\mathbf{g}_n(\hat{\boldsymbol{\theta}}_n)\| \xrightarrow{p} 0$, CQFD.

Proof of Proposition 2

We check that all the conditions of theorem 7.2 of Newey and McFadden (1994, p. 2186) are verified. We first need to check that $\mathbf{g}_n(\hat{\boldsymbol{\theta}}_n)$ verifies an "asymptotic first-order condition," $\mathbf{g}_n(\hat{\boldsymbol{\theta}}_n)'\hat{\mathbf{S}}_n^{-1}\mathbf{g}_n(\hat{\boldsymbol{\theta}}_n) \le \inf_{\boldsymbol{\theta}\in\Theta}\mathbf{g}_n(\boldsymbol{\theta})'\hat{\mathbf{S}}_n^{-1}\mathbf{g}_n(\boldsymbol{\theta}) + o_p(n^{-1})$. For this, it suffices to have $\sqrt{n}\|\mathbf{g}_n(\hat{\boldsymbol{\theta}}_n)\| \xrightarrow{p} 0$, which is shown to hold in Lemma A.3. Note that we also have $\hat{\mathbf{S}}_n \xrightarrow{p} \mathbf{S}$ with $\mathbf{S}$ nonsingular, so that $\hat{\mathbf{S}}_n^{-1} \xrightarrow{p} \mathbf{S}^{-1}$. Moreover, $\mathbf{S}^{-1}$ is positive definite. We now check conditions 7.2(i)–7.2(v). By definition, $\boldsymbol{\theta}^*$ is a solution to $\mathbf{g}_0(\boldsymbol{\theta}^*) = \mathbf{0}$, which shows that 7.2(i) holds. To show that 7.2(ii) holds, note that $\mathbf{g}$ can be written as $\mathbf{g}(\boldsymbol{\theta}; Y_{t+1}, \mathbf{W}_t^*) = [\alpha - H(\boldsymbol{\theta}'\hat{\mathbf{q}}_{m,t} - Y_{t+1})]\mathbf{W}_t^*$, where $H(\cdot)$ is the Heaviside function, that is, $H(x) = 1$ if $x > 0$ and 0 if $x < 0$. The "gradient" of $\mathbf{g}(\cdot)$ is the function $\boldsymbol{\Delta}(\cdot)$ with

$$\boldsymbol{\Delta}(\boldsymbol{\theta}; Y_{t+1}, \mathbf{W}_t^*) \equiv -\delta(\boldsymbol{\theta}'\hat{\mathbf{q}}_{m,t} - Y_{t+1})\mathbf{W}_t^*\hat{\mathbf{q}}_{m,t}', \qquad \text{(A.3)}$$

where $\delta(\cdot)$ represents the Dirac function, that is, $\delta(x) = 0$, if $x \ne 0$ and $\int_{\mathbb{R}}\delta(x)\,dx = 1$. Note that $\delta(\cdot)$ is the derivative of $H(\cdot)$, so that we have $|H(x + \varepsilon) - H(x) - \varepsilon\delta(x)| = o(|\varepsilon|)$ for all $x \in \mathbb{R}$. We now show that $\boldsymbol{\Delta}$ is indeed a "gradient" of $\mathbf{g}$ in a neighborhood of $\boldsymbol{\theta}^*$, in the sense that $\|\mathbf{g}(\boldsymbol{\theta}; Y_{t+1}, \mathbf{W}_t^*) -$

$\mathbf{g}(\boldsymbol{\theta}^*; Y_{t+1}, \mathbf{W}_t^*) - \boldsymbol{\Delta}(\boldsymbol{\theta}^*; Y_{t+1}, \mathbf{W}_t^*)(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\| = o_p(\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|)$. Let

$$r(\boldsymbol{\theta}^*; Y_{t+1}, \mathbf{W}_t^*) \equiv \big\|\mathbf{g}(\boldsymbol{\theta}; Y_{t+1}, \mathbf{W}_t^*) - \mathbf{g}(\boldsymbol{\theta}^*; Y_{t+1}, \mathbf{W}_t^*) - \boldsymbol{\Delta}(\boldsymbol{\theta}^*; Y_{t+1}, \mathbf{W}_t^*)(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\big\| / \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|.$$

To simplify the notation, we drop the reference to t and let $X \equiv \boldsymbol{\theta}^{*\prime}\hat{\mathbf{q}}_{m,t} - Y_{t+1}$ and $\varepsilon \equiv (\boldsymbol{\theta} - \boldsymbol{\theta}^*)'\hat{\mathbf{q}}_{m,t}$. Thus

$$\begin{aligned} &r(\boldsymbol{\theta}^*; Y_{t+1}, \mathbf{W}_t^*) \\ &\quad = \|\mathbf{W}_t^*\| \cdot |H(X+\varepsilon) - H(X) - \varepsilon \cdot \delta(X)| / \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \\ &\quad \le \|\mathbf{W}_t^*\| \cdot \|\hat{\mathbf{q}}_{m,t}\| \cdot |H(X+\varepsilon) - H(X) - \varepsilon \cdot \delta(X)| / |\varepsilon|, \qquad \text{a.s.-}P, \end{aligned}$$

where we used the fact that $|\varepsilon| \le \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \cdot \|\hat{\mathbf{q}}_{m,t}\|$. Let $A_t \equiv \|\mathbf{W}_t^*\| \cdot \|\hat{\mathbf{q}}_{m,t}\|$. By the Cauchy–Schwartz inequality, we have $E(A_t^2) \le [E\|\mathbf{W}_t^*\|^4]^{1/2}[E\|\hat{\mathbf{q}}_{m,t}\|^4]^{1/2}$, so that assumptions (e) and (f) imply that $E(A_t^2) < \infty$. We now use the finiteness of the second moment of A_t to construct an upper bound for $P(r(\boldsymbol{\theta}^*; Y_{t+1}, \mathbf{W}_t^*) > \epsilon)$. For any $\eta > 0$ and any $\epsilon > 0$, let $A \equiv [2E(A_t^2)/\eta]^{1/2} < \infty$ and $\tilde{\epsilon} \equiv \epsilon/A > 0$; we then have

$$\begin{aligned} &P\big(r(\boldsymbol{\theta}^*; Y_{t+1}, \mathbf{W}_t^*) > \epsilon\big) \\ &\quad \le P\big(A_t \cdot |H(X+\varepsilon) - H(X) - \varepsilon \cdot \delta(X)| / |\varepsilon| > \epsilon\big) \\ &\quad \le P\big(A_t \cdot |H(X+\varepsilon) - H(X) - \varepsilon \cdot \delta(X)| / |\varepsilon| > \epsilon | A_t \le A\big) \\ &\qquad \times P(A_t \le A) \\ &\qquad + P\big(A_t \cdot |H(X+\varepsilon) - H(X) - \varepsilon \cdot \delta(X)| / |\varepsilon| > \epsilon | A_t > A\big) \\ &\qquad \times P(A_t > A) \\ &\quad \le P\big(|H(X+\varepsilon) - H(X) - \varepsilon \cdot \delta(X)| / |\varepsilon| > \epsilon/A\big) \\ &\qquad + P(A_t > A), \end{aligned}$$

so that, by Chebyshev's inequality,

$$\begin{aligned} &P\big(r(\boldsymbol{\theta}^*; Y_{t+1}, \mathbf{W}_t^*) > \epsilon\big) \\ &\quad \le P\big(|H(X+\varepsilon) - H(X) - \varepsilon \cdot \delta(X)| / |\varepsilon| > \tilde{\epsilon}\big) + 1/A^2 \cdot E(A_t^2) \\ &\quad \le P\big(|H(X+\varepsilon) - H(X) - \varepsilon \cdot \delta(X)| / |\varepsilon| > \tilde{\epsilon}\big) + \eta/2. \end{aligned}$$

Because Dirac delta function is the derivative of Heaviside function, we know that given $\tilde{\epsilon} > 0$ and $\eta' \equiv \eta/3 > 0$, there exist some $e > 0$ such that $|\varepsilon| < e$ implies that $P(|H(X + \varepsilon) - H(X) - \varepsilon \cdot \delta(X)| / |\varepsilon| > \tilde{\epsilon}) < \eta/3$. Further, recall that $\varepsilon \equiv (\boldsymbol{\theta} - \boldsymbol{\theta}^*)'\hat{\mathbf{q}}_{m,t}$, so that for any $e > 0$, there exist some $\rho > 0$ such that $\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| < \rho$ implies that $|\varepsilon| < e$ and thus implies that $P(|H(X+\varepsilon) - H(X) - \varepsilon \cdot \delta(X)| / |\varepsilon| > \tilde{\epsilon}) < \eta/3$. For any $\eta > 0$ and any $\epsilon > 0$, we have found $\rho > 0$ such that $\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| < \rho$ implies that $P(r(\boldsymbol{\theta}^*; Y_{t+1}, W_t^*) > \epsilon) < \eta$; that is, we have shown that $P(\lim_{\boldsymbol{\theta} \to \boldsymbol{\theta}^*} r(\boldsymbol{\theta}^*; Y_{t+1}, \mathbf{W}_t^*) = 0) = 1$. Therefore, we can say that $\mathbf{g}_0(\boldsymbol{\theta}) = E[\mathbf{g}(\boldsymbol{\theta}; Y_{t+1}, \mathbf{W}_t^*)]$ is differentiable at $\boldsymbol{\theta}^*$ with derivative $\boldsymbol{\gamma} \equiv E[\boldsymbol{\Delta}(\boldsymbol{\theta}^*; Y_{t+1}, \mathbf{W}_t^*)]$. Using the expression in (A.3), note that

$$\begin{aligned} \boldsymbol{\gamma} &= E[-\delta(\boldsymbol{\theta}^{*\prime}\hat{\mathbf{q}}_{m,t} - Y_{t+1})\mathbf{W}_t^*\hat{\mathbf{q}}_{m,t}'] \\ &= E[E_t[-\delta(\boldsymbol{\theta}^{*\prime}\hat{\mathbf{q}}_{m,t} - Y_{t+1})]\mathbf{W}_t^*\hat{\mathbf{q}}_{m,t}'] \\ &= -E[f_t(\boldsymbol{\theta}^{*\prime}\hat{\mathbf{q}}_{m,t})\mathbf{W}_t^*\hat{\mathbf{q}}_{m,t}'], \end{aligned}$$

where $f_t(\cdot)$ is the density of Y_{t+1} conditional on the information set $\mathcal{F}_t$. We now show that $\boldsymbol{\gamma}'\mathbf{S}^{-1}\boldsymbol{\gamma}$ is nonsingular. As before, consider the quadratic form $\boldsymbol{\zeta}'\boldsymbol{\gamma}'\mathbf{S}^{-1}\boldsymbol{\gamma}\boldsymbol{\zeta}$, where $\boldsymbol{\zeta} \in \mathbb{R}^2$. We have $\boldsymbol{\zeta}'\boldsymbol{\gamma}'\mathbf{S}^{-1}\boldsymbol{\gamma}\boldsymbol{\zeta} = 0$ if and only if $\boldsymbol{\gamma}\boldsymbol{\zeta} = \mathbf{0} \in \mathbb{R}^h$ because, as shown in Proposition 1, $\mathbf{S}^{-1}$ is positive definite. On the other hand $\boldsymbol{\gamma}\boldsymbol{\zeta} = \mathbf{0}$ if and only if $E[f_t(\boldsymbol{\theta}^{*\prime}\hat{\mathbf{q}}_{m,t})\mathbf{W}_t^*\hat{\mathbf{q}}_{m,t}'\boldsymbol{\zeta}] = 0$. Given that f_t is assumed to be strictly positive, this last equality holds only if $\hat{\mathbf{q}}_{m,t}'\boldsymbol{\zeta} = 0$, a.s.-$P$. Because, by assumption (b), $\hat{q}_{i,m,t} \ne 0$, a.s.-P, the previous condition can hold only if $\boldsymbol{\zeta} = \mathbf{0}$. Hence $\boldsymbol{\gamma}'\mathbf{S}^{-1}\boldsymbol{\gamma}$ is nonsingular.

Condition 7.2(iii) is trivially satisfied by imposing that Θ be compact. To show that 7.2(iv) holds, that is, that $\sqrt{n}\mathbf{g}_n(\boldsymbol{\theta}^*) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{S})$, we use a central limit theorem (CLT) for martingale difference sequences (e.g., corollary 5.26 in White 2001, p. 135). Recall from (8) that $\{\mathbf{g}(\boldsymbol{\theta}^*; Y_{t+1}, \mathbf{W}_t^*), \mathcal{F}_t\}$ is a martingale difference sequence. Also, as shown in the proof of Proposition 1, $\hat{\mathbf{S}}_n(\boldsymbol{\theta}^*) \xrightarrow{p} \mathbf{S}$. To apply the CLT provided in corollary 5.26 of White (2001), we need to show that $E\|\mathbf{g}(\boldsymbol{\theta}^*; Y_{t+1}, \mathbf{W}_t^*)\|^{2+\delta} < \infty$ for some $\delta > 0$. We have $E\|\mathbf{g}(\boldsymbol{\theta}; Y_{t+1}, \mathbf{W}_t^*)\|^{2+\delta} \le E\|\mathbf{W}_t^*\|^{2+\delta} \le \max\{1, E\|\mathbf{W}_t^*\|^{2r+\delta}\}$, where $r > 2$, so that by assumption (e), $E\|\mathbf{g}(\boldsymbol{\theta}; Y_{t+1}, \mathbf{W}_t^*)\|^{2+\delta} < \infty$ for some $\delta > 0$. We can therefore use corollary 5.26 of White (2001) to show that $\sqrt{n}\mathbf{g}_n(\boldsymbol{\theta}^*) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{S})$.

Finally, Andrews (1994) has shown that the stochastic equicontinuity condition 7.2(v) holds for moment functions such as $\mathbf{g}(\boldsymbol{\theta}; Y_{t+1}, \mathbf{W}_t^*)$. We can now apply the results of theorem 7.2 of Newey and McFadden (1994) to show that $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) \xrightarrow{d} \mathcal{N}(\mathbf{0}, (\boldsymbol{\gamma}'\mathbf{S}^{-1}\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{S}^{-1}\boldsymbol{\gamma}(\boldsymbol{\gamma}'\mathbf{S}^{-1}\boldsymbol{\gamma})^{-1})$, that is, $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) \xrightarrow{d} \mathcal{N}(\mathbf{0}, (\boldsymbol{\gamma}'\mathbf{S}^{-1}\boldsymbol{\gamma})^{-1})$, CQFD.

Proof of Theorem 1

From Proposition 2, it follows that $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$, with $\boldsymbol{\Omega} = (\boldsymbol{\gamma}'\mathbf{S}^{-1}\boldsymbol{\gamma})^{-1}$ nonsingular. Given a consistent estimate $\hat{\boldsymbol{\Omega}}_n$ of $\boldsymbol{\Omega}$, we have that $n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*)'\hat{\boldsymbol{\Omega}}_n^{-1}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) \xrightarrow{d} \chi_2^2$ (e.g., thm. 4.30 of White 2001), from which (a) and (b) follow.

Proof of Lemma 2

We need to show that $\operatorname{plim}_{n\to\infty}(\lim_{\tau\to 0}\hat{\boldsymbol{\gamma}}_{n,\tau}) = \boldsymbol{\gamma}$, where $\boldsymbol{\gamma} = \lim_{\tau\to 0}\boldsymbol{\gamma}_\tau$ and $\boldsymbol{\gamma}_\tau \equiv \operatorname{plim}_{n\to\infty}\hat{\boldsymbol{\gamma}}_{n,\tau}$, that is, $\boldsymbol{\gamma}_\tau = E\{\frac{1}{\tau}\exp[(Y_{t+1} - \boldsymbol{\theta}^{*\prime}\hat{\mathbf{q}}_{m,t})/\tau]\mathbb{1}(Y_{t+1} - \boldsymbol{\theta}^{*\prime}\hat{\mathbf{q}}_{m,t} < 0)\mathbf{W}_t^*\hat{\mathbf{q}}_{m,t}'\}$. To change the order of the two limits, we need to make sure that $\hat{\boldsymbol{\gamma}}_{n,\tau}$ converges to $\boldsymbol{\gamma}_\tau$ uniformly in τ in some neighborhood of 0 (see, e.g., thm. 2.13.18 in Schwartz 1997). First, note that for any $u \in \mathbb{R}^+$, the function $\tau \mapsto \frac{1}{\tau}\exp(u/\tau)\mathbb{1}(u < 0)$ is identically equal to 0, whereas for any $u \in \mathbb{R}_-^*$, it is convex in τ on $]0, a[$, where $a \equiv -u(1 - 1/\sqrt{2}) > 0$. Hence if $\hat{\boldsymbol{\gamma}}_{n,\tau}$ converges to $\boldsymbol{\gamma}_\tau$ pointwise, then the convergence is also uniform in τ provided that τ remains in a neighborhood around 0. More formally, let $U_{t+1} \equiv Y_{t+1} - \hat{\boldsymbol{\theta}}_n'\hat{\mathbf{q}}_{m,t}$, for any t, $m \le t \le T-1$. We then have $\forall\,\varepsilon > 0$ and $\forall\,\eta > 0$,

$$\begin{aligned} &P\Big(\sup_{\tau\in A}|\hat{\boldsymbol{\gamma}}_{n,\tau} - \boldsymbol{\gamma}_\tau| \ge \varepsilon\Big) \\ &\quad \le P\Bigg(\sup_{\tau\in A}\frac{1}{\tau}\sup_{m\le t\le T-1}\big|\exp(U_{t+1}/\tau)\mathbb{1}(U_{t+1} < 0)\mathbf{W}_t^*\hat{\mathbf{q}}_{m,t}' \end{aligned}$$

$$- E[\exp(U_{t+1}/\tau)\mathbb{1}(U_{t+1} < 0)\mathbf{W}_t^*\hat{\mathbf{q}}'_{m,t}]\big| \geq \varepsilon\Big)$$

$$= P\Big(\sup_{\tau\in A}\frac{1}{\tau}\sup_{m\leq t\leq T-1}\big|\exp(U_{t+1}/\tau)\mathbb{1}(U_{t+1} < 0)\mathbf{W}_t^*\hat{\mathbf{q}}'_{m,t}$$

$$- E[\exp(U_{t+1}/\tau)\mathbb{1}(U_{t+1} < 0)\mathbf{W}_t^*\hat{\mathbf{q}}'_{m,t}]\big| \geq \varepsilon,$$

$$\sup_{m\leq t\leq T-1, U_{t+1}<0} U_{t+1} \leq -\eta\Big)$$

$$+ P\Big(\sup_{m\leq t\leq T-1, U_{t+1}<0} U_{t+1} > -\eta\Big).$$

Now $P(\sup_{m\leq t\leq T-1, U_{t+1}<0} U_{t+1} > -\eta) \leq \prod_{m\leq t\leq T-1, U_{t+1}<0}\eta \times f_t(\hat{\boldsymbol{\theta}}'_n\hat{\mathbf{q}}_{m,t})$, where f_t is the conditional density of Y_{t+1}. Hence under assumption (g) when f_t is bounded above by some constant C, we have $P(\sup_{m\leq t\leq T-1, U_{t+1}<0} U_{t+1} > \eta) \leq \eta^{n_-}C^{n_-}$, where $n_- \equiv \#\{U_{t+1}, m \leq t \leq T-1 : U_{t+1} < 0\}$. For any $\nu > 0$, then let η be such that $\eta^{n_-}C^{n_-} = \nu/2$, that is, $\eta = C^{-1}\exp[n_-^{-1}\ln(\nu/2)] > 0$. Let $A \equiv]0, \eta(1-1/\sqrt{2})[$; then for any t, $m \leq t \leq T-1$, such that $U_{t+1} < 0$, we have $-U_{t+1} \geq \eta$, and hence the function $\tau \mapsto \frac{1}{\tau}\exp(u/\tau)\mathbb{1}(u < 0)$ is convex in τ on A. This implies that for any t, $m \leq t \leq T-1$, the convergence of $\tau^{-1}\exp(U_{t+1}/\tau)\mathbb{1}(U_{t+1} < 0)\mathbf{W}_t^*\hat{\mathbf{q}}'_{m,t}$ to its expected value is uniform on A, so that the first term of the right side converges to 0 and can be made arbitrarily small, say smaller than $\nu/2$. To resume, we can make each term of the foregoing inequality smaller than $\nu/2$, provided that we have the pointwise convergence of the series $\{\tau^{-1}\exp(U_{t+1}/\tau)\mathbb{1}(U_{t+1} < 0)\mathbf{W}_t^*\hat{\mathbf{q}}'_{m,t}\}$. The latter is easy to show by an LLN for α-mixing sequences. Specifically, we check that all the assumptions of corollary 3.48 of White (2001) hold. First, note that for $r > 2$, we have $-r/(r-1) > -r/(r-2)$, so that under assumption (c), the sequence $\{\tau^{-1}\exp(U_{t+1}/\tau)\mathbb{1}(U_{t+1} < 0)\mathbf{W}_t^*\hat{\mathbf{q}}'_{m,t}\}$ is α-mixing with α of size $-r/(r-1)$. We now need to show that for some $\tilde{\delta} > 0$, we have $E\|\tau^{-1}\exp(U_{t+1}/\tau)\mathbb{1}(U_{t+1} < 0)\mathbf{W}_t^*\hat{\mathbf{q}}'_{m,t}\|^{r+\tilde{\delta}} < \infty$. Note that we have $\|\tau^{-1}\exp(U_{t+1}/\tau)\mathbb{1}(U_{t+1} < 0)\mathbf{W}_t^*\hat{\mathbf{q}}'_{m,t}\| \leq \tau^{-1}\|\mathbf{W}_t^*\hat{\mathbf{q}}'_{m,t}\|$, a.s.-$P$. Hence the Cauchy–Schwartz inequality, together with assumptions (e) and (f), then ensure that $E\|\tau^{-1}\exp(U_{t+1}/\tau)\mathbb{1}(U_{t+1} < 0)\mathbf{W}_t^*\hat{\mathbf{q}}'_{m,t}\|^{r+\tilde{\delta}} < \infty$. Applying corollary 3.48 of White (2001) and using the consistency of $\hat{\boldsymbol{\theta}}_n$ for $\boldsymbol{\theta}^*$ then gives $\hat{\boldsymbol{\gamma}}_{n,\tau} \xrightarrow{p} \boldsymbol{\gamma}_\tau = E\{\frac{1}{\tau}\exp[(Y_{t+1} - \boldsymbol{\theta}^{*\prime}\hat{\mathbf{q}}_{m,t})/\tau] \times \mathbb{1}(Y_{t+1} - \boldsymbol{\theta}^{*\prime}\hat{\mathbf{q}}_{m,t} < 0)\mathbf{W}_t^*\hat{\mathbf{q}}'_{m,t}\}$. The remainder of the proof is straightforward. We know that $\lim_{\tau\to 0}\hat{\boldsymbol{\gamma}}_{n,\tau} \xrightarrow{p} \boldsymbol{\gamma}$ and $\hat{\mathbf{S}}_n \xrightarrow{p} \mathbf{S}$ with $\mathbf{S}$ nonsingular, so $\lim_{\tau\to 0}\hat{\boldsymbol{\gamma}}_{n,\tau}\hat{\mathbf{S}}_n^{-1}\hat{\boldsymbol{\gamma}}_{n,\tau} = (\lim_{\tau\to 0}\hat{\boldsymbol{\gamma}}_{n,\tau}) \times \hat{\mathbf{S}}_n^{-1}(\lim_{\tau\to 0}\hat{\boldsymbol{\gamma}}_{n,\tau}) \xrightarrow{p} \boldsymbol{\gamma}\mathbf{S}^{-1}\boldsymbol{\gamma}$ nonsingular, yielding $\hat{\boldsymbol{\Omega}}_n = \lim_{\tau\to 0}(\hat{\boldsymbol{\gamma}}_{n,\tau}\hat{\mathbf{S}}_n^{-1}\hat{\boldsymbol{\gamma}}_{n,\tau})^{-1} \xrightarrow{p} (\boldsymbol{\gamma}\mathbf{S}^{-1}\boldsymbol{\gamma})^{-1} = \boldsymbol{\Omega}$, CQFD.

[Received July 2002. Revised September 2004.]

REFERENCES

Andrews, D. W. K. (1994), "Empirical Process Methods in Econometrics," in *Handbook of Econometrics*, Vol. 4, eds. R. F. Engle and D. L. McFadden, New York: North-Holland, pp. 2247–2294.

Barone-Adesi, G., Bourgoin, F., and Giannopoulos, K. (1998), "Don't Look Back," *Risk*, 11, 100–104.

Bates, J. M., and Granger, C. W. J. (1969), "The Combination of Forecasts," *Operational Research Quarterly*, 20, 451–468.

Battacharya, P. K., and Gangopadhyay, A. K. (1990), "Kernel and a Nearest-Neighbor Estimation of a Conditional Quantile," *The Annals of Statistics*, 18, 1400–1415.

Bierens, H. J., and Ginther, D. (2001), "Integrated Conditional Moment Testing of Quantile Regression Models," *Empirical Economics*, 26, 307–324.

Bracewell, R. N. (2000), *The Fourier Transform and Its Applications* (3rd ed.), New York: McGraw-Hill.

Chernozhukov, V., and Umantsev, L. (2001), "Conditional Value-at-Risk: Aspects of Modeling and Estimation," *Empirical Economics,* 26, 271–292.

Christoffersen, P. (1998), "Evaluating Interval Forecasts," *International Economic Review*, 39, 841–862.

Christoffersen, P., Hahn, J., and Inoue, A. (2001), "Testing and Comparing Value-at-Risk Measures," *Journal of Empirical Finance*, 8, 325–342.

Clements, M. P., and Hendry, D. F. (1998), *Forecasting Economic Time Series*, Cambridge: Cambridge, U.K.: University Press.

Corradi, V., and Swanson, N. R. (2002), "A Consistent Test for Nonlinear Out-of-Sample Predictive Accuracy," *Journal of Econometrics*, 110, 353–381.

——— (2004), "Bootstrap Procedures for Recursive Estimation Schemes With Applications to Forecast Model Selection," unpublished manuscript, Rutgers University.

Danielsson, J., and de Vries, C. (1997), "Tail Index and Quantile Estimation With Very High Frequency Data," *Journal of Empirical Finance*, 4, 241–257.

Diebold, F. X. (1989), "Forecast Combination and Encompassing: Reconciling Two Divergent Literatures," *International Journal of Forecasting*, 5, 589–592.

Diebold, F. X., and Mariano, R. S. (1995), "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13, 253–263.

Diebold, F. X., Schuermann, T., and Stroughair, J. (1998), "Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management," in *Advances in Computational Finance*, eds. A.-P. N. Refenes, J. D. Moody, and A. N. Burgess, Amsterdam: Kluwer Academic, pp. 3–12.

Duffie, D., and Pan, J. (1997), "An Overview of Value at Risk," *Journal of Derivatives*, 4, 7–49.

Elliott, G., and Timmermann, A. (2004), "Optimal Forecast Combinations Under General Loss Functions and Forecast Error Distributions," *Journal of Econometrics*, 122, 47–80.

Embrechts, P., Resnick, S., and Samorodnitsky, G. (1999), "Extreme Value Theory as a Risk Management Tool," *North American Actuarial Journal*, 3, 30–41.

Engle, R. F., and Manganelli, S. (2004), "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles," *Journal of Business & Economic Statistics*, 22, 367–381.

Giacomini, R., and White, H. (2003), "Tests of Conditional Predictive Ability," unpublished manuscript, University of California, San Diego.

Granger, C. W. J. (1989), "Combining Forecasts—Twenty Years Later," *Journal of Forecasting*, 8, 167–173.

Granger, C. W. J., and Ramanathan, R. (1984), "Improved Methods of Combining Forecasts," *Journal of Forecasting*, 3, 197–204.

Hansen, L. P. (1982), "Large-Sample Properties of Generalized Method-of-Moments Estimators," *Econometrica*, 50, 1029–1054.

Hendry, D. F., and Richard, J. F. (1982), "On the Formulation of Empirical Models in Dynamic Econometrics," *Journal of Econometrics*, 20, 3–33.

Kitamura, Y., and Stutzer, M. (1997), "An Information-Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica*, 65, 861–874.

Koenker, R. W., and Bassett, G. W. (1978), "Regression Quantiles," *Econometrica*, 46, 33–50.

Koenker, R., and Zhao, Q. (1996), "Conditional Quantile Estimation and Inference for ARCH Models," *Econometric Theory*, 12, 793–813.

Komunjer, I. (2005), "Quasi-Maximum Likelihood Estimation for Conditional Quantiles," *Journal of Econometrics*, to appear.

McCracken, M. (2000), "Robust Out-of-Sample Inference," *Journal of Econometrics*, 99, 195–223.

McNeil, A. J., and Frey, R. (2000), "Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach," *Journal of Empirical Finance*, 7, 271–300.

Mizon, G. E., and Richard, J. F. (1986), "The Encompassing Principle and Its Application to Testing Non-Nested Hypotheses," *Econometrica*, 54, 657–678.

JP Morgan, (1996), *RiskMetrics* (4th ed.), New York: JP Morgan.

Newey, W. K., and McFadden, D. L. (1994), "Large-Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, Vol. 4, eds. R. F. Engle and D. L. McFadden, New York: North-Holland, pp. 2113–2247.

Newey, W. K., and West, K. D. (1987), "A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Schwartz, L. (1997), *Analyse*, Vol.1, Paris: Hermann.

Stock, J. H., and Watson, M. W. (1999), "A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series," in *Cointegration, Causality and Forecasting*, eds. R. F. Engle and H. White, Oxford, U.K.: Oxford University Press, pp. 1–44.

——— (2003), "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature*, 41, 788–829.

Taylor, J. (1999), "A Quantile Regression Approach to Estimating the Distribution of Multi-Period Returns," *Journal of Derivatives*, 7, 64–78.

Taylor, J., and Bunn, D. W. (1998), "Combining Forecast Quantiles Using Quantile Regression: Investigating the Derived Weights, Estimator Bias and Imposing Constraints," *Journal of Applied Statistics*, 25, 193–206.

West, K. D. (1996), "Asymptotic Inference About Predictive Ability," *Econometrica*, 64, 1067–1084.

——— (2001), "Encompassing Tests When No Model Is Encompassing," *Journal of Econometrics*, 105, 287–308.

White, H. (ed.) (1992), "Nonparametric Estimation of Conditional Quantiles Using Neural Networks," in *Artificial Neural Networks: Approximation and Learning Theory*, Oxford, U.K.: Blackwell, pp. 191–205.

——— (2000), "A Reality Check for Data Snooping," *Econometrica*, 68, 1097–1126.

——— (2001), *Asymptotic Theory for Econometricians*, San Diego: Academic Press.

Zheng, J. X. (1998), "A Consistent Nonparametric Test of Parametric Regression Models Under Conditional Quantile Restrictions," *Econometric Theory*, 14, 123–138.